# nature

**CANDIDA GENOMICS**

**ANTARCTIC GLACIATION**

**NANOMATERIALS**
Semiconductor crystals that don't blink

# SEEING IS ACHIEVING
**New miracles from microscopes**

# Abstractions

**SECOND AUTHOR**
Formation of the Antarctic Ice Sheet was set in motion some 34 million years ago, when atmospheric carbon dioxide levels dropped drastically, causing rapid global cooling. However, it remains unclear how early glaciation events developed into a continent-sized sheet of ice. On page 690, an international team of scientists use ice-penetrating radar to pinpoint the centre of the ice sheet's growth. Beneath 3,000 metres of ice, they describe a picturesque landscape, akin to that of the Alps, that has remained unchanged for at least 14 million years. Martin Siegert, a glaciologist at the University of Edinburgh, UK, tells *Nature* more.

**How did this collaboration come about?**
At a meeting in France in 2002, I met lead author Sun Bo of the Shanghai-based Polar Research Institute of China, and he then visited me for a year at the University of Bristol's Glaciology Centre, where I was director. We gave him access to the largest existing data set of Antarctic ice-penetrating radar, collected by aircraft in the 1970s. He spent two years planning a field survey of the region at the centre of the ice sheet, named Dome A, that overlies the Gamburtsev Mountains. He wanted to map the subglacial landscape in this remote location and, in so doing, document the early events of glaciation, as a way of advancing China's ambitious Antarctic research programme.

**Was it difficult to survey the region?**
Yes. The conditions are so extreme that few people have ever visited the site. It is a desperately cold place at high altitude, set more than 4 kilometres above sea level. Sun Bo and his colleagues drove snowmobile sleds carrying radar survey equipment from the ice sheet's edge to its centre to survey an area 30 kilometres square around Dome A. This was no small achievement.

**What can the Alp-like topography tell us?**
That it was carved during two earlier smaller-scale glacial phases and then preserved under a large sheet of ice. The lack of erosion that had occurred since the ice sheet formed allowed Sun Bo and I to determine that these early glaciation features would have formed at about 3 °C. Because the records we have contain no evidence of Antarctic temperatures being near that level during the past 14 million years, we know that the ice sheet is at least that old.

**Does this work shed light on current climate change?**
It provides an understanding of how little the Antarctic Ice Sheet has changed over 14 million years, during which carbon dioxide levels have been relatively stable. ∎

# MAKING THE PAPER

## Dan Tawfik & Nobuhiko Tokuriki

### Flooding the cell with chaperones speeds up test-tube evolution.

**Dan Tawfik (left) and Nobuhiko Tokuriki.**

For much of Nobuhiko 'Nobu' Tokuriki's tenure as a postdoc in Dan Tawfik's lab at the Weizmann Institute of Science in Rehovot, Israel, he was the only Japanese researcher in residence — earning him visits from the Japanese ambassador whenever he was in town. But Tokuriki did not mind his unique status. "I wanted to see a foreign society and culture," he says. "And I wanted to study protein evolution." For that, Tawfik's lab was the place to be.
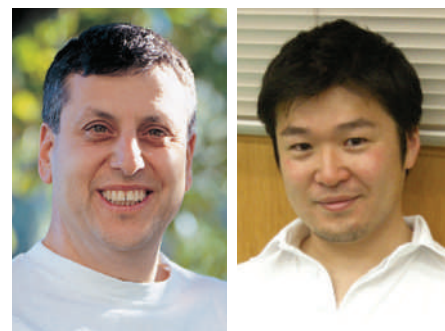
In nature, proteins evolve new functions rapidly and efficiently. Researchers have long tried to reproduce cellular conditions in the test tube to better understand the process of protein evolution, as well as to search for new or improved enzymes with industrial or medical applications. But such methods, typically referred to as 'directed evolution', have had limited success, resulting in processes that are sluggish compared with those in nature. With Tokuriki's help, Tawfik may have found a way to greatly improve the efficiency of directed evolution by boosting the activities of the cell's quality-control officers — the chaperone proteins.

The method has its origin in work that came to Tawfik's attention about seven years ago. At the time, most scientists thought that the degree to which a mutation improved or degraded a protein's activity was what drove its positive or negative selection.

But in 2002, work by Brian Shoichet and his colleagues, then at Northwestern University in Chicago, Illinois, seemed to suggest that many mutations never actually see the light of day. Shoichet's group proposed that certain mutations would gravely affect the way that newly made proteins are folded into three-dimensional shapes, affecting their stability. Thus, the proteins would be quickly whisked away as unstable misfolded 'trash'. And even though some such mutations might be 'adaptive', or of improved function, they would never be detected. Later research indicated that this destabilizing effect of mutation is one of the major hurdles to protein evolution.

After Shoichet's work was published, Tawfik recalls, "I read it again and again. But it took time for its implications to penetrate into the field." Once the concept sank in, however, it was a short jump for Tawfik to the idea that chaperone proteins, which normally rescue misfolded proteins, might be able to rescue some mutant proteins and accelerate directed protein evolution.

As this idea was taking shape, Tokuriki joined Tawfik's lab, keen to work on the project. "Although it is far from perfect, protein evolution is one of the few areas where we are able to get clear biophysical and biochemical clues as to how and why evolution in the test tube happens," says Tokuriki. "This is why I chose to study this and not organism evolution."

Over the course of three-and-a-half years, Tokuriki performed directed evolution experiments on selected enzymes from the bacterium *Escherichia coli*, both in the presence of large amounts of a chaperone protein called GroEL/ GroES and in its absence. This chaperone normally isolates unstable, misfolded proteins in the cell and gives them space to try to refold properly so that they can function normally. The work required each of hundreds of mutant proteins to be characterized over several bacterial generations. "Nobu can do the work of a dozen," Tawfik says.

The gruelling work revealed that, overall, GroEL/GroES rescued about one-third of the adaptive mutant proteins that, without the chaperone's aid, would have been too unstable to be viable. As a result, it allowed twice as many mutations to accumulate in proteins.

> **"It's clear there are certain mutations you would never see without this method."**

Next, the duo tested the ability of GroEL/GroES to speed up the evolution of a new, divergent function in a specific enzyme. In the chaperone's presence, directed evolution produced at least twice as many adapted variations than in its absence, and these were at least ten times more active and specific than those that evolved without the chaperone (see page 668).

Tawfik says that even though the idea was something "we thought would be stupid not to try", he was still shocked by how well it worked. Using chaperones in directed evolution will almost certainly be embraced by those trying to produce more powerful enzymes for industrial or therapeutic applications, says Tawfik, "because it's clear there are certain adaptive mutations that you would never see without this method". ∎

# Microscopic marvels

**Microscopes are changing the face of biology. Researchers should innovate and collaborate if they want to be part of the new vision.**

Watching molecular-scale events unfold in a living cell can be an inspiring experience. The inner workings of the nucleus, the shuttling of cellular cargo, the passage of messages through a membrane — seeing this tumultuous activity up close can fire the scientific imagination in a way that abstract data from genetic sequences or chemical analyses can never quite equal.

This helps to explain why microscopes are such essential tools in science, and why scientists' desire to see more is driving rapid innovation in the field. Five microscopes representative of these innovations are featured in this issue, starting on page 629. They range from 'super-resolution' devices that use light to reveal details once visible only with electron microscopes, to an electron microscope that can peer into thick samples once accessible only to optical ones.

Few of these new imaging technologies come cheaply, however. One new-generation light microscope can easily cost US$500,000 or more, not counting the staff and training required to use it. There is no room for error on an instrument so sensitive that the slightest vibration or misstep in the experimental set-up creates artefacts.

As a result, biologists may have to get used to sharing their instruments. Research groups often prefer to buy their own microscopes so that lab members can use them whenever they want. But to stay at the forefront of image-led science, they will increasingly have to work with their departments, universities and funding agencies to create shared microscopy facilities staffed by specialists.

Getting engineers into the mix will be a good idea, and computational biologists will be a must. Automation, computer processing and image reconstruction are now central to microscopy and imaging. This means that there is increasing separation between the sample placed in the microscope and what scientists actually see, and that there are numerous points at which inaccuracies can inadvertently be introduced. The only way to interpret the images the system spits out is to understand (and explain alongside the final results) precisely what went into their creation. Not that this problem is new: scientists

have struggled with how to interpret microscope images since the instrument's earliest days (see page 642).

As well as sharing microscopes, cell and molecular biologists will need to share the instruments' output. A good way to do that would be through a central repository or framework for biological images modelled on existing data repositories such as GenBank. Such a resource would not only help to make the images universally available — using widely accepted data standards — but could be a driver for the development of analysis techniques that extract deeper meaning from existing images. A database launched last year by *The Journal of Cell Biology* (http://jcb-dataviewer. rupress.org) is already a start towards this goal. Such experiments deserve the community's support and participation.

Meanwhile, some labs are exploring an alternative to the fat-price-tag rule: instead of buying one machine for $500,000, buy 50,000 microscopes for $10 apiece. By creating microscopes that are small, cheap and even disposable, these researchers hope to accelerate the development of microscopy into a high-throughput, automated procedure that can quickly collect data on living cells as systematic changes are made to one gene, protein or receptor at time. The era of systems microscopy could soon be here.

> **"The era of systems microscopy could soon be here."**

The wonderful thing about thinking visually is that it is so easy to think big. And that's exactly what researchers should do as they move ahead. By thinking about what they really want to see, they will help to devise microscopes — as well as computation, labelling and sample-preparation techniques — that make that vision possible. Thinking big might lead to microscopy that allows individual molecules to be tracked across thousands of cells in real time in living tissue; that can watch a single cell over the months or years that lapse from birth to death; or that can map the intricate form of every neuron in the brain across multiple species. The deeper biologists look, the more they will find there is to see. ∎

# Stem-cell clarity

**The draft NIH guidelines on stem-cell research are a good first step, but some revision is needed.**

The proposed guidelines on federal funding for stem-cell research issued in April by the US National Institutes of Health (NIH) are a welcome effort to assert ethical and regulatory leadership over this field — especially given the vacuum in oversight left by the previous US administration. Yet concerns aired by the scientific community during the public comment period that closed last week have underscored the need for the NIH to revise the

guidelines to allow the responsible progress of research.

For example, the NIH has acted admirably in setting forth nine strict informed-consent provisions regarding the source material for stem-cell lines that are eligible for federal funding. As currently written, however, the provisions would probably exclude funding for most embryonic stem cells now in use, because the cells were derived from leftover embryos at fertility clinics under rules less stringent than the ones now called for. In particular, one provision requires that embryo donors affirm that they are donating "without any restriction or direction" regarding the patients who may benefit from the research. Few of the consent forms currently in use by fertility clinics ask for that affirmation. Also absent from many forms is the stipulation that the donors "would not receive financial or any other

benefits" from commercial development of the research.

It is doubtful that the guidelines are intended to bar future NIH funding from stem-cell lines that are currently eligible — or, indeed, from the hundreds of lines that are now in use but are not among the score of US-approved lines. That would contradict US President Barack Obama's intention, stated on 9 March, to "expand NIH support for the exploration of human stem cell research". The NIH should explicitly state that the informed-consent provisions apply only to newly created lines. All previously eligible lines should continue to be eligible, and existing non-eligible lines should become so — provided that the latter were created in accordance with guidelines issued by the US National Academies or the International Society for Stem Cell Research.

A much less clear-cut issue is whether federal funding should support work on lines created from sources other than leftover embryos. The NIH's draft guidelines currently exclude support for research on lines created through somatic-cell nuclear transfer (SCNT), also known as therapeutic cloning. And they prohibit funding for lines created through the generation of an embryo from an unfertilized egg cell.

Some investigators have protested this provision of the guidelines, arguing that the NIH should not cut off any avenues of research. Their contention is somewhat hypothetical, however, because no one has yet shown conclusive evidence that SCNT can successfully create a human embryo. Regardless of this, the ethical issues involved are extremely sensitive. Polls consistently show that a majority of the American public is willing to pay for research on stem cells derived from embryos that would be discarded otherwise. But it is not clear that a majority would support the use of taxpayers' money to study stem cells from embryos created and destroyed for research purposes alone. So unless the scientists arguing for federal funding of research on SCNT-derived stem cells can make a much stronger case, by spelling out the specific situations in which the research might be warranted and explaining how they will ensure proper oversight of the work, the NIH's proposed exclusion should stand.

At the same time, however, the NIH should affirm that it will revisit its draft guidelines as the science progresses. The past decade shows us that basing research policy on arbitrary cut-off dates does not serve science or the public interest well. ∎

# Power vacuum

**The US president's delay in naming an NIH director is symptomatic of a widespread problem.**

As *Nature* went to press, US president Barack Obama had still not nominated a director for the National Institutes of Health (NIH), the $30.3-billion agency that is the world's largest funder of biomedical research. Regardless of when a director is named, the delay is already too long: Obama took office 136 days ago. In the interim, the NIH has been forced to navigate multiple sensitive issues under temporary leadership, including decisions about how to spend a massive $10 billion in economic stimulus money; the drafting of guidelines for expanded federal funding of human embryonic stem-cell research (see above); and the launch of a proposal to tighten conflict-of-interest reporting requirements for its extramural investigators.

Moreover, the problem goes well beyond the NIH. The installation of senior agency leaders, most of whom have to be nominated by the president and confirmed by a majority vote of the US Senate, seems to go ever more slowly with each passing administration.

Granted, the Obama White House has been trying. As of Monday the Senate had confirmed 145 people for a total of 373 jobs that need filling, which is not too different from the pace set by the incoming Bush administration in 2001. But that still leaves only 4 of 21 Senate-confirmable posts filled at the Department of Health and Human Services (HHS). Among those HHS posts still waiting for a permanent occupant — in addition to the NIH directorship — are the assistant secretary for health, the top public health adviser to the new HHS secretary, Kathleen Sebelius; the assistant secretary for planning and evaluation, her top policy adviser; and the assistant secretary for preparedness and response, her top adviser on bioterrorism and other public-health emergencies.

Even appointments that don't require Senate confirmation have been slow to materialize. Richard Besser, the acting director at the Centers for Disease Control and Prevention, was left to deal with the emergence of swine flu — Thomas Frieden, Obama's permanent appointee for the job, won't begin work until 8 June.

Other science agencies are in similar straits. At the National Oceanic and Atmospheric Administration, Jane Lubchenco was confirmed as administrator in mid-March, but she still lacks a chief scientist and two assistant administrators, one for atmosphere and one for oceans. No nominees are in sight. At NASA, the two top posts went unfilled until 23 May, when Obama finally nominated former space-shuttle astronaut Charles Bolden as administrator and Lori Garver as his deputy. But no Senate hearings have yet been scheduled for them. And at the National Science Foundation, there is still no nomination for a deputy director — a key strategic post that encompasses the duties of chief operating officer.

This dilatory pace is partly a result of Obama's promise to run a squeaky-clean administration staffed by officials not beholden to lobbyists or other moneyed interests. But it is mostly the result of the Senate confirmation process, which in recent decades has become increasingly obsessed with savaging nominees over even the most minor slip-ups on taxes or nannies. To minimize the chances of embarrassment, the administration now requires Senate-confirmable nominees to fill out lengthy vetting questionnaires. These are so onerous that many feel compelled to spend their own money hiring accountants and lawyers to help fill them in. Other qualified people simply refuse to go through such an ordeal, and take themselves out of contention.

The administration and the Senate together must find a way to restore common sense to this process. Given the challenges faced by the United States, ranging from nuclear proliferation to climate change and potential pandemics, its government needs to recruit the best minds it can find — without subjecting them to a protracted, politically motivated vetting process that does nothing to solve real problems. ∎

# RESEARCH HIGHLIGHTS

## Pretty please

*Anim. Behav.* doi:10.1016/j.anbehav.2009.03.11 (2009)

Many young animals beg for food from their elders. But, eventually, the pleading stops or the charity dries up. Joah Madden, at the University of Cambridge, UK, and his team looked to find the biological triggers that put an end to begging behaviour by studying free-ranging meerkats (*Suricata suricatta*) of the Kalahari Desert in South Africa over an 18-month period.

The group analysed the begging calls of meerkat pups aged between 40 and 60 days — the peak of their begging behaviour — and compared them with the calls of the same individuals aged 100–120 days. Experimental playback to adults revealed that lower-pitched juvenile calls reaped fewer rewards than the pleading of pups.

R. WITTEK/GETTY

### EVOLUTION
## Carnivore claim quashed

*Proc. Natl Acad. Sci. USA* doi:10.1073_pnas.0901780106 (2009)

A plausible hypothesis states that socialization in mammals puts them under selective pressure to evolve larger brains.

John Finarelli of the University of Michigan in Ann Arbor and John Flynn of the American Museum of Natural History in New York compared relative brain sizes for the group Carnivora, which includes cats, dogs, bears and weasels. They looked at 289 examples, 125 of them extinct, and found that brain-size changes through evolutionary time do not correlate well with the development of sociality.

The authors suggest that the 'social brain hypothesis' falls apart when looking at carnivore groups, extinct and otherwise, beyond modern canids (wolves, jackals, foxes and the like).

### ASTRONOMY
## A big head start

*Astrophys. J.* **697,** 1493–1511 (2009)

Many astronomers have thought that most of the large, mature galaxies in the Universe took a long time to get that way, as smaller galaxies burst onto the scene and merged into larger ones over billions of years. But Daniel Stark of the University of Cambridge, UK, and his colleagues have found that the early Universe may have been a more active galactic crucible.

Using Hubble Space Telescope data, the researchers identified more than 3,000 galaxies that formed up to about 12 billion years ago, when the Universe was only 1.7 billion years old. Many of these galaxies were already big, suggesting that these earliest epochs of the Universe were very busy times, and probably responsible for the formation of a significant component of the Universe's mature, massive galaxies.

### PALAEONTOLOGY
## Bone study bugbear

*Proc R. Soc. B* doi:10.1098/rspb.2009.0563 (2009)

Palaeontologists may need to find a better way to predict the integrity of DNA in ancient remains.

Processes to extract DNA from fossils can cause considerable specimen damage (as pictured, right). To avoid needless destruction, researchers have tended to screen samples first, measuring the extent of aspartic acid racemization — the chemical conversion of the amino acid to a different form over time — to estimate protein integrity. The belief was that if proteins remained intact, so too did DNA. Now Matthew Collins at the University of York, UK, and his colleagues reveal that this is not the case.

They studied 91 specimens of tooth and bone, evaluating how protein analysis compared with rates of DNA amplification success. Racemization offered little information on the quality of the DNA in a fossil.

### GENETICS
## Setting the biological clock

*Nature Genet.* **41,** 724–728; 734–738; 729–733 (2009)

A series of studies has tracked possible genetic influences on when a woman's reproductive lifespan begins and ends.

Chunyan He of the Harvard School of Public Health and her colleagues scanned the genomes of more than 17,000 women, looking for genetic sequences associated with age at menarche — the start of the first menstrual cycle. They found a series of genetic markers associated with the onset of sexual maturity, including several clustered in and near a gene called *LIN28B*, and additional markers associated with the timing of menopause. Another project, by Patrick Sulem and Kari Stefansson of deCODE Genetics in Reykjavik and their collaborators, also found a link between *LIN28B* and the onset of puberty. Meanwhile, a third study from Ken Ong and Ruth Loos of Addenbrooke's Hospital in Cambridge, UK, and their colleagues reports a particular form of the gene that is associated with earlier menarche and breast development in girls, and earlier voice-breaking in boys.
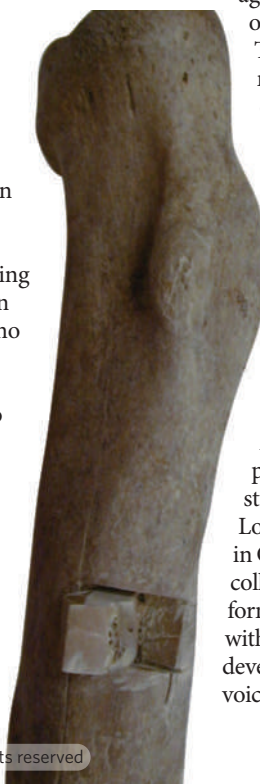
B. SHAPIRO & A. COOPER

## NEUROBIOLOGY
# Squeaking in tongues

*Cell* **137**, 961–971 (2009)

Mice carrying a humanized version of a gene considered key to the human capacity for speech show subtle changes in their brains and in the way they vocalize.

Wolfgang Enard of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, and his colleagues substituted two amino acids in the mouse Foxp2 protein to generate the humanized version. The animals remained generally healthy, but calls made by isolated pups had a different structure from those of normal mouse pups.

The authors also found structural, neurochemical and neurophysiological changes to neurons in a brain circuit associated with speech in humans. They say it could represent the first investigation, in an animal model, of amino-acid substitutions that might be relevant to human evolution.
**For a movie about this work, see http://bit.ly/tLtFa.**

## CHEMISTRY
# Toxic toadstools

*Nature Chem. Biol.* doi:10.1038/nchembio.179 (2009)

A form of mushroom poisoning that leads to often-fatal breakdown of skeletal muscle tissue has been claiming unlucky victims in Japan since at least the 1950s. However, the toxin responsible has eluded mycologists owing to both its instability and confusion about the classification of the *Russula* mushrooms that it is associated with, say Kimiko Hashimoto of Kyoto Pharmaceutical University in Japan and her colleagues.

They found that *Russula subnigricans* (pictured below) from Kyoto were toxic when fed to mice, whereas similar mushrooms from Miyagi prefecture were not. In addition, they identified the toxic compound responsible for the poisoning: cycloprop-2-ene carboxylic acid.

Extrapolating from data on mice, the researchers say that a serving of two to three mushroom caps (less than 200 grams) would kill most humans.

## GENETICS
# Filling mouse holes

*PLoS Biol.* **7**, e1000112 (2009)

Humans and mice have slightly less in common than was previously thought, reveals the international Mouse Genome Sequencing Consortium.

The team produced a more complete *Mus musculus* genome by sequencing DNA fragments and linking them back to a physical map of the genome. Earlier efforts used an efficient approach known as whole-genome sequence and assembly, which does



not require a physical map. This left many gaps owing to difficulty in assembling some regions, such as those containing repetitive DNA. Now researchers have filled more than 175,000 gaps and uncovered more than 139 million bases of overlooked DNA. Moreover, they identified more than 2,000 genes the sequences of which were missed or wrongly assembled before, more than half of which do not have an analogue in the human genome.

The results underscore the importance of complete genome sequences, and should lend more power to comparative studies in human biology, the authors suggest.

## QUANTUM PHYSICS
# Attack of the giant neutrinos

*Phys. Rev. Lett.* **102**, 201303 (2009)

Neutrinos are vanishingly small particles, but a theoretical analysis of neutrinos left over from the Big Bang shows that they could stretch across billions of light years.

Like all fundamental particles, the state and position of neutrinos can be described with a quantum-mechanical 'wavefunction'. George Fuller and Chad Kishimoto of the University of California, San Diego, looked at the wavefunctions of neutrinos left over from the very beginning of the Universe. They found that these waves could extend for billions of light years. If true, that would mean that the individual neutrinos are effectively spread out over the same cosmic distances.

Disturbances could cause the wavefunction to collapse, and Fuller and Kishimoto postulate that gravitational effects or space-time itself might make the little particles pop into a single location.

K. HASHIMOTO

---

# JOURNAL CLUB

**Lucas N. Joppa**
**Nicholas School of the Environment, Duke University, Durham, North Carolina**

**An ecologist calls for a citizen-science 'Wiki'.**

Where do species occur and why? What happens to ecological communities when species are removed or when alien species invade? And how will the answers shift as climates change? These questions span huge spatial and temporal scales, and involve millions of species. By contrast, ecological field studies are generally of short duration, include few species and cover small areas. This means that getting data for the big questions is a tall order — impossible without harnessing a deeper reserve of people power.

Citizen science — in which qualified scientists oversee volunteers — is not new. The Audubon Society's Christmas Bird Count has run for 108 years and mobilizes about 60,000 volunteers across 1 million square kilometres of North America who count about 58 million individual birds annually. Other citizen-science projects are under way around the world.

Dirk Schmeller at the Helmholtz Centre for Environmental Research in Leipzig, Germany, and his colleagues analysed 395 citizen-science projects across five European countries, involving more than 46,000 participants (D. Schmeller *Conserv. Biol.* **23**, 307–316; 2008). Volunteers donated more than 148,000 person-days per year, a figure inconceivable using professional scientists alone. Schmeller *et al.* found that volunteer-gathered data are reliable and unbiased, with data quality determined less by 'volunteer' status, and more by survey design and methodology.

The rapid increase in citizen-science data sets can revolutionize what we know about the natural world. Wikipedia has shown that the public is willing to donate time, talent and knowledge, given a sufficient platform. Biodiversity data lack such a platform for input, integration, mapping and dissemination. This is a deficiency that the environmental community should address.

Discuss this paper at **http://blogs.nature.com/nature/journalclub**

# NEWS

# Mouse patent sparks 'uncivil' spat

## Academic institutes lock horns in legal action over mutant mice.

The Jackson Laboratory, a non-profit genetics research centre in Bar Harbor, Maine, is embroiled in its first ever patent dispute — with another non-profit research institute.

"It is almost unheard of for one academic institute to sue another over patent infringement," says the Jackson Lab's corporate lawyer, David Einhorn, who described the dispute on 21 May at a meeting in Rome on sharing data and resources for functional genomics. "The affair is completely uncivil."

The suit was filed by the Central Institute for Experimental Animals (CIEA) in Kawasaki, Japan, which last week reported the world's first transgenic primate able to pass a foreign gene to its offspring (E. Sasaki *et al. Nature* **459,** 523–527; 2009). In 2006 the CIEA was granted a US patent on a strain of immunodeficient mice called NOG, which it began to make commercially available in the United States earlier this year. The institute has also trademarked the strain.

The CIEA's lawsuit claims that the Jackson Lab is infringing the patent, a claim the laboratory denies. Among the more than 4,000 mouse mutant strains that the Jackson Lab maintains and distributes to the worldwide academic community is a mouse similar to NOG, called NSG. The Jackson Lab developed the NSG strain and, since 2006, has made it available to more than 640 research groups worldwide. It does not patent.

The CIEA says it is after recognition, not profits. "We would like scientists around the world to recognize the innovation involved in the creation of the NOG mouse," says Hideo Maeno, general manager in the institute's strategic management department.

The institute distributes NOG mice under quite restrictive terms: purchasers cannot breed or cross-breed from them, and there are extensive 'reach-through' rights that control successive work done with the mice.

Immunodeficient mice are widely used in the study of diseases such as cancer or diabetes because they can be 'humanized' by having human tissue or human transgenes transplanted into them. As their immune system is suppressed, the mice tend not to reject the foreign tissue.

In the 1980s, an immunodeficient mouse strain called CB17-SCID was generated by a spontaneous mutation. In the 1990s, this SCID mutation was crossed first by scientists at the Jackson Lab, and later at the CIEA, with their respective strains of another, non-obese diabetic (NOD) mouse.

The Jackson Lab and the CIEA then each independently crossed their NOD-SCID strains with mice whose gene for a key immune signalling molecule, IL2Rγ, was either partially or completely knocked out. The NOD-SCID IL2Rγ-deficient mice — abbreviated as NOG by the CIEA and as NSG by the Jackson Lab — are particularly good at accepting grafted tissue.
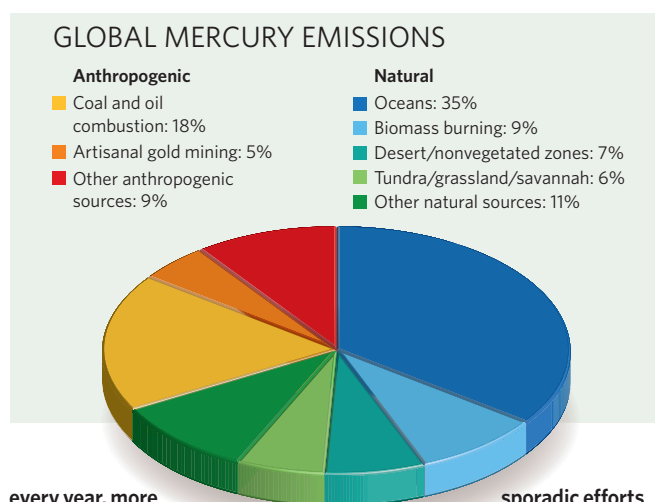
The Jackson Lab denies that it has infringed any patent or trademark rights. Einhorn notes that the long-established NOD and SCID strains used to create the mice were not identical anyway, because of natural genetic drift accumulated over many generations in the separate locations.

# Funding struggle for mercury monitoring

**Nicola Pirrone may need all the help he can get next week. In the hallways of a conference in Guiyang, China, Pirrone — the director of Italy's CNR-Institute for Atmospheric Pollution Research — will be trying to rustle up support for a global network to monitor mercury pollution.**

Such a network would underlie a United Nations Environment Programme (UNEP) treaty to control mercury emissions, which negotiators plan to forge by 2013. So far, countries from Mexico to South Africa to Japan have expressed interest in setting up a monitoring system. But to turn interest into reality, researchers are facing a complex task on what may be shoestring budgets.

Anthropogenic sources emit about 2,500 tonnes of mercury



**GLOBAL MERCURY EMISSIONS**

**Anthropogenic**
- Coal and oil combustion: 18%
- Artisanal gold mining: 5%
- Other anthropogenic sources: 9%

**Natural**
- Oceans: 35%
- Biomass burning: 9%
- Desert/nonvegetated zones: 7%
- Tundra/grassland/savannah: 6%
- Other natural sources: 11%

every year, more than half of which comes from fossil-fuel power plants (see chart). But so far, global monitoring endeavours have been relatively uncoordinated; hundreds of

sporadic efforts can include one-time samplings from a ship cruise or aeroplane flight.

The United States and Canada do have networked and standardized

stations that can monitor mercury deposition at a high enough precision over the long term to be useful in a treaty. These include sophisticated monitors made by Tekran Instruments Corporation in Knoxville, Tennessee, which can collect mercury from both rainfall and dry air and determine its chemical form.

US researchers have been working to combine data from about a dozen Tekran-equipped sites, along with other longer-term studies, into a national monitoring effort called MercNet that could serve as a starting point for a global network. Some initial high-quality worldwide data are starting to become available, for instance from Taiwan's Changbai Mountain (Q. Wan *et al. Environ. Res.* 109, **201–206; 2009**).

But expanding the number of

**ELECTRODES SPARK NEURON GROWTH**
Deep-brain stimulation could help make memories.
www.nature.com/news
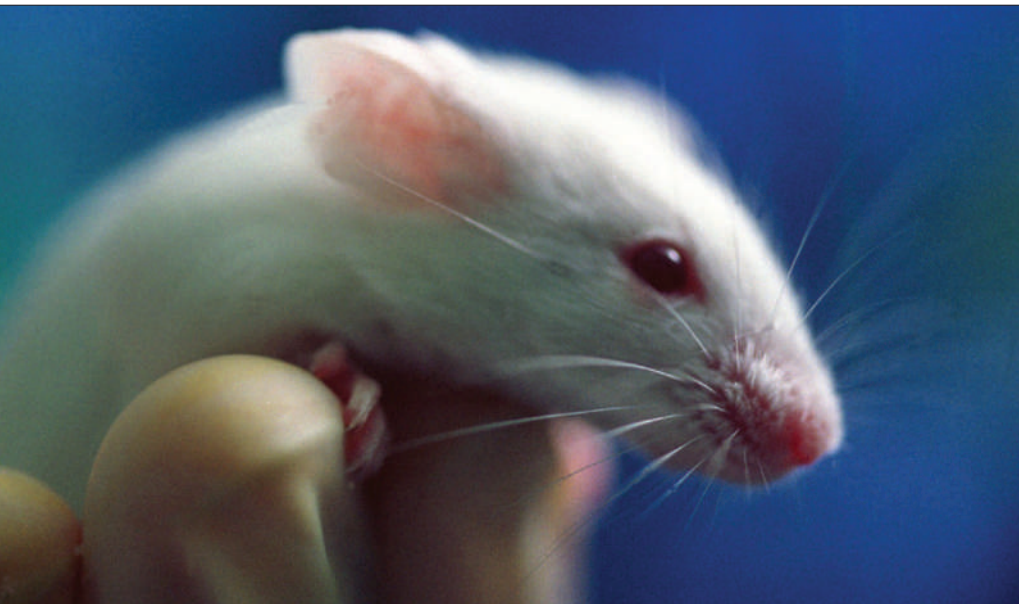


Handle with care: more than 4,000 strains of mice are produced at the Jackson Laboratory.

that the CIEA infringed a patent it filed on the knockout mouse used to breed both the NOG and NSG strains. The NIH patents biological inventions to prevent others from patenting and then restricting use.

Jackson Lab scientist Leonard Shultz, who has worked for 35 years on developing humanized mice including the NSG strain, echoes the concerns of many in the community when he says he does not understand the CIEA's motivation. The CIEA says it cannot comment concretely on such matters, because of the ongoing litigation.

On face value, the CIEA does not seem like an aggressively modern, lawyer-wielding institute. It is more old-world, with its rundown buildings tucked into a modest residential area of the working-class city of Kawasaki.

It is still directed by 87-year-old Tatsuji Nomura, the infectious-disease researcher who founded it in 1952 because he was concerned about the variable quality of research animals. It is operated by a foundation funded by some 30 Japanese pharmaceutical and biomedical companies.

"Science is essentially a global enterprise and national profit is not part of our mindset," insists Maeno. "We are absolutely not trying to slow the pace of science." ∎

**Alison Abbott**
**Additional reporting by David Cyranoski**

Einhorn is particularly upset that the CIEA did not contact the lab about its concern before filing the case in December in a US district court in northern California. "The first I heard was three days later when I got a call from a Californian lawyer who had spotted the case and was offering us his services," he says.

Maeno did not respond directly to questions about why his institute didn't contact the Jackson Lab before filing suit. But, he says, "we have enjoyed a close relationship with the Jackson Lab in the past and we hope to continue that relationship in the future."

The National Institutes of Health, which funds the Jackson Lab repository, has asked the lab to bring a counterclaim on its behalf, saying

---

sites globally will be expensive. A top-of-the-line Tekran monitor costs about US$100,000, not including operational costs. On top of that, mercury researchers would like to see a host of other measurements, including regularly scheduled aeroplane monitoring high in the troposphere to see how chemical reactions change airborne mercury; research cruises to measure what is happening in the ocean–air boundary layer; and more data on mercury levels in the oceans themselves.

One money-saving option might be to piggyback mercury measurements on other sites that already collect atmospheric data, says Robert Mason, an environmental chemist at the University of Connecticut in Groton. Mason is advising researchers who are working to set up a network in China, Japan and Korea, including expanding existing stations.

But some researchers worry that the focus on air monitoring may mean that other parts of the

environment get overlooked. "There has to be a biological component to this," says David Evers of the BioDiversity Research Institute in Gorham, Maine. That could mean, for instance, identifying 'hot spots' that receive airborne mercury and turn it into a form that could be taken up by fish and then humans.

Pierrette Blanchard, an atmospheric chemist at Environment Canada, notes that even fish in seemingly pristine lakes in national forests in the United States and Canada can contain high levels of mercury.

In particular, Evers wants to see estuarine and marine sampling in the Amazon basin, where small-scale gold mining releases mercury into environments where fish are caught for food. He and his colleagues have also proposed a marine and estuaries network for monitoring the US Atlantic coast.

And Evers and others have been advising Senator Susan Collins (Republican, Maine) on a bill that

would provide funding from the US Environmental Protection Agency for a national mercury network to include air, water, soils and animals. Collins has tried several times before; in 2007, for instance, she requested that $18 million be designated, but the effort died at the committee stage. Collins hopes to reintroduce the bill this summer.

The costs of an international monitoring system are unknown, and will depend on how much various countries are willing to put up for equipment and staffing. That may be a point for discussion when negotiations for the UNEP treaty begin in earnest early next year. The talks have been boosted by the United States, which had long called for voluntary-only mercury monitoring but in February indicated it would support an

international regulatory structure. "What we need to do is to inform the negotiations," says a US state department official who requested anonymity because of the talks.

Bettina Hitzfeld, a UNEP negotiator for the Swiss Federal Office for the Environment, says that treaty discussions need to start moving ahead even in the absence of a complete monitoring system. "We cannot afford to wait," she says, "for a mercury network to be up and running." ∎

Naomi Lubick

**H1N1 update**

As confirmed by the World Health Organization, as of 1 June:
Cases: 17,410, in 62 countries
Deaths: 115, in 4 countries

UNITED STATES: Cases have now been documented in all 50 states.

For *Nature*'s swine flu coverage, see www.nature.com/swineflu.

# SPECIAL REPORT

# Europe's parliamentary priorities

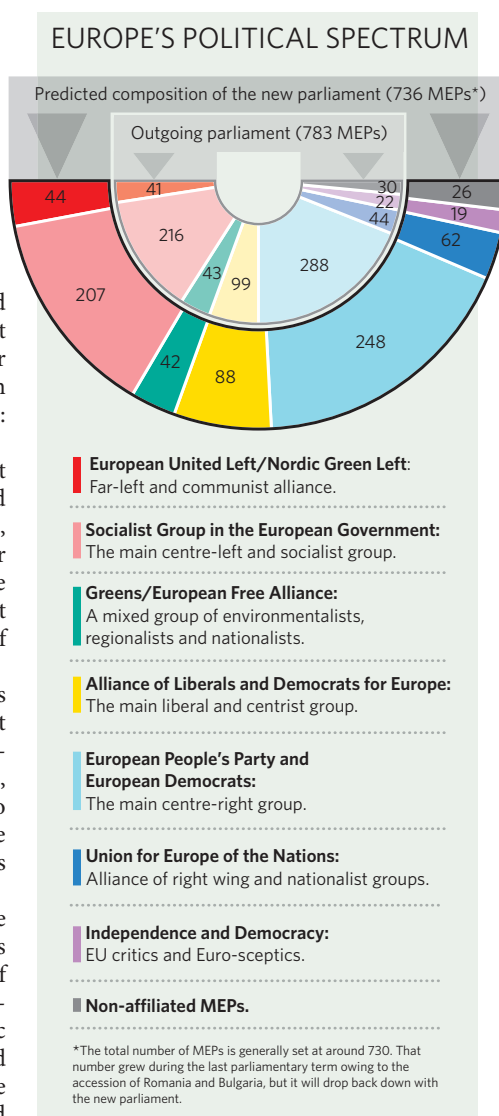This week's Europe-wide elections will affect researchers more than they realize, reports **Alison Abbott**.

Chris Davies, a former British Member of Parliament who has been a Member of the European Parliament (MEP) for the last ten years, is convinced that "you can do more for environment legislation in Brussels through parliamentary committee work than you can in any national parliament". His view is shared by many MEPs across parties, political groups, nations and political themes. The European Parliament has an intrinsic dynamism, they say, and although it has no tax-raising powers, it can get things done. And more than any of the other institutions of the European Union (EU), it can allow elected individuals to make a difference: personal enthusiasm counts.

"The European Parliament is very important in developing the European Research Area and influencing its budget," says Philippe Busquin, a former European Research Commissioner and an MEP (Socialist Group, Belgium) in the outgoing parliament. "So it is very important to elect MEPs who are attentive to the issues of science and technology."

In the past decade, the parliament has demonstrated that it can have a direct effect on European (and global) science and technology policies. Against all the odds, for example, it managed to find financing for the Galileo satellite navigation system in 2007 when the ministers of member states insisted there was no money to spare.

The parliament also has a direct effect on the working lives of scientists and the research paths available to them. In May, in the final days of the last legislature, MEPs saved proposed legislation governing the use of animals in scientific experiments from going down a path that would have outlawed key areas of medical research (see *Nature* **459,** 139; 2009). But they are also behind some of the bizarre complexity that scientists find themselves faced with when applying for grants through the EU Framework Programme — such as the requirement to explain the gender relevance of research on single-celled organisms, as on every other topic.

In the elections on 4–7 June this year, 380 million electors in the union's 27 countries will be able to choose the parliament's 736 members, although if the turnout is similar to that seen in 2004, well under half will bother to vote. The new MEPs will take their seats in July and serve for five years; they will thus be deeply involved in the development



## EUROPE'S POLITICAL SPECTRUM

Predicted composition of the new parliament (736 MEPs*)

Outgoing parliament (783 MEPs)

| **European United Left/Nordic Green Left**: Far-left and communist alliance. |
| **Socialist Group in the European Government:** The main centre-left and socialist group. |
| **Greens/European Free Alliance:** A mixed group of environmentalists, regionalists and nationalists. |
| **Alliance of Liberals and Democrats for Europe:** The main liberal and centrist group. |
| **European People's Party and European Democrats:** The main centre-right group. |
| **Union for Europe of the Nations:** Alliance of right wing and nationalist groups. |
| **Independence and Democracy:** EU critics and Euro-sceptics. |
| **Non-affiliated MEPs.** |

*The total number of MEPs is generally set at around 730. That number grew during the last parliamentary term owing to the accession of Romania and Bulgaria, but it will drop back down with the new parliament.

of the multi-billion-euro Eighth Framework Programme (FP8), which should launch in 2014. The parliament will also have to consider whether to allow the fledgling European Research Council to become an independent agency or to remain attached to the European Commission, with all its attendant bureaucracy, as part of FP8. With the right MEPs in place, says Busquin, lobbying for high research budgets is more likely to be successful.

The European Parliament has considerable, but not absolute, power in such issues, as it is just one of the EU's three key institutions (see

'What does the European Parliament do?'). Yet its influence has long been underestimated, for two main reasons. In the past, its decision-making power was much more limited than it is now, and perceptions have been slow to catch up with the gradual increase of power brought about by successive treaties. That power will grow again if the 2007 Lisbon Treaty is ratified later this year by the last few member states that have not yet signed up. Under that treaty, the parliament's realm of codecision would be extended to new areas, including agriculture and fisheries policies and the allocation of structural funds — subsidies to poor EU regions that can be used to help build up research infrastructure.

Second, the complexity of parliament means that many people are flummoxed by its decision-making processes. The 200 or so national political parties represented in the 2004–09 parliament sat in seven political groups (see 'Europe's political spectrum'). Each group is a relatively loose coalition without the sort of party whip familiar in national parliaments; no group has consensus on all policies. Sensitive ethical issues, such as whether human embryonic stem cells should be used in EU-funded research, for example, are voted on according to conscience rather than party membership. Voting on other issues, such as those relating to development of particular technologies, is sometimes governed more by national allegiance than ideology, with members keeping an eye out for domestic industries.

## Expert input

Most of the parliament's work is done in expert committees, which examine proposed legislation. MEPs divide themselves among the committees, and each of the political groups has at least one expert policy adviser (who is not an elected politician). Work on the animal-testing directive, for example, was led by the agriculture committee, with input from the research and environment committees. At the first reading, the MEPs and their expert advisers were quick to see major problems in the commission's draft and insert corrective amendments; these were then approved in the plenary session in May.
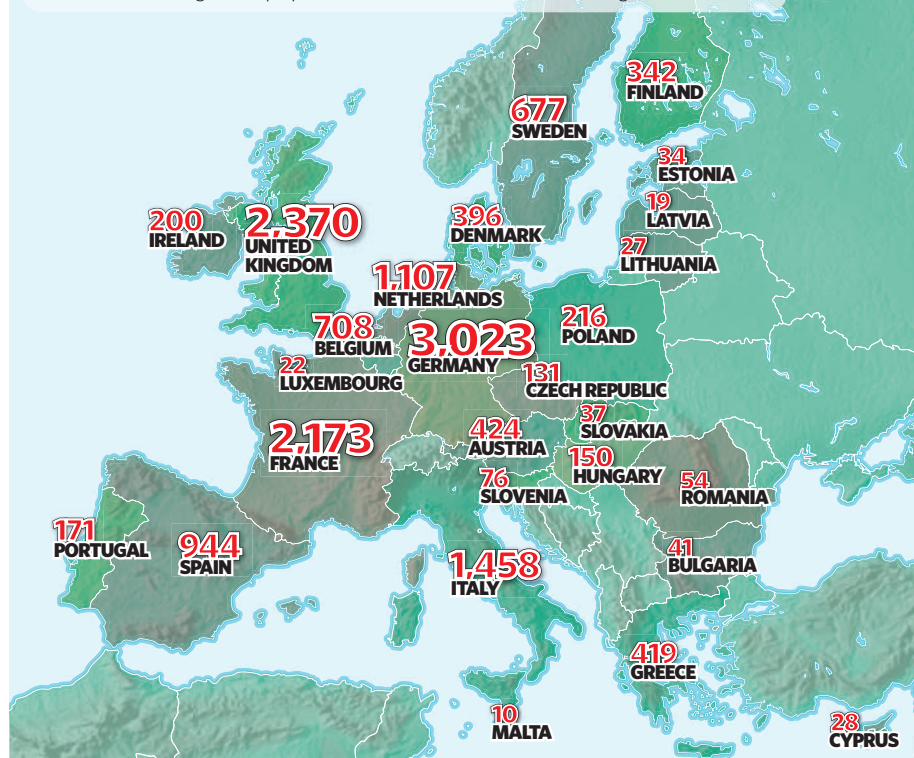
This was far from the largest of the science-related issues the last parliament had to deal with. The giant €50.5-billion FP7 programme was finally approved in 2006. In the process it was seen by 8 committees and MEPs had to

**EXPLAINING DARK ENERGY**
'Chameleon' particle changes mass to match surroundings.
www.nature.com/news

## SHARING THE WEALTH

Between 2002 and 2006, Europe's Sixth Framework Programme for Research and Technological Development disbursed €16.7 billion between participating states (figures below in millions of euros). The Seventh Framework Programme, covering the period 2007 to 2013, is likely to see member states taking similar proportions of more than €50 billion in funding.



- **342** FINLAND
- **677** SWEDEN
- **34** ESTONIA
- **19** LATVIA
- **27** LITHUANIA
- **200** IRELAND
- **2,370** UNITED KINGDOM
- **396** DENMARK
- **1,107** NETHERLANDS
- **216** POLAND
- **708** BELGIUM
- **3,023** GERMANY
- **22** LUXEMBOURG
- **131** CZECH REPUBLIC
- **2,173** FRANCE
- **424** AUSTRIA
- **37** SLOVAKIA
- **76** SLOVENIA
- **150** HUNGARY
- **54** ROMANIA
- **171** PORTUGAL
- **944** SPAIN
- **1,458** ITALY
- **41** BULGARIA
- **419** GREECE
- **10** MALTA
- **28** CYPRUS

council or the parliament were persuaded, and the institute was cut down to a pilot scheme of large research networks. "But we introduced the word 'Innovation' into the title," says Reino Paasilinna, a Finnish Social Democrat who sits in the Socialist group. Paasilinna sees his mission as to "mainstream innovation and research into each bit of legislation".

### European stand-off

With the exception of FP8, the 2009–2014 parliament is likely to have fewer major pieces of science-related legislation to get to grips with, but there will be plenty of fine-tuning and updating, particularly to bring regulations covering climate and energy in line with evolving international agreements, as well as to oversee the planned evaluations of the newer institutions such as the EIT and the European Research Council. It could also find itself revisiting EU legislation allowing cultivation of genetically modified crops, subject to safety control, an issue on which some individual states have reached a stand-off against the union.

This should leave the parliament with the energy to deal with two important, difficult to resolve, issues for European research. One concerns the support of infrastructure — such as the planned European Mouse Mutant Archive — a project that may be too large for an individual nation, but is not strictly in the EU's legal remit. The second concerns the mobility of researchers, who often find it difficult to move smoothly from institution to institution because of the differing national employment laws, or because of visa requirements. The mobility issue is a hard one to crack because it involves so many different areas of legislation. "But it is absolutely key for the future of Europe," says Paasilinna.

"The great thing about the European Parliament is that we can usually find consensus," adds Davies. "We are not beholden to our national governments — the separation between the European institutions allows us to explore common ground across parties rather than having to knee-jerk reject what the opposition suggests." ∎

sift through more than 1,500 proposed amendments. Nearly all legislation passes through the equal-opportunities committee, which brings in gender issues wherever it sees an opportunity; this is where the demand for justification of research projects in terms of gender were spawned. In 2006, the parliament dealt with the Registration, Evaluation, Authorisation and Restriction of Chemical substances (REACH) regulation, the largest piece of legislature — and one of the most contentious — ever handled by the European Union. This directive went to 10 committees and MEPs had to consider 2,400 amendments. And in 2008, the parliament approved a package of climate and energy laws to cut greenhouse-gas emissions and promote renewable energies according to legally binding targets. MEPs insisted on a package without wiggle room. "It was parliament that insisted, against the will of the council of ministers, that the targets remain binding," says Umberto Guidoni (Group of the European United Left, Italy), a former astronaut.

Stand-offs between the parliament and the council also ensured that smaller research

> **"It is very important to elect MEPs who are attentive to the issues of science and technology."**

initiatives continued to promote innovation in small and medium-sized enterprises and determined the final — very modest — shape of the European Institute of Innovation and Technology (EIT), says MEP Jorgo Chatzimarkakis, a member of the German Free Democratic party who sits in the liberal Alliance of Liberals and Democrats for Europe group. Commission president José Manuel Barroso had envisioned the institute as a mighty competitor for the Massachusetts Institute of Technology in Cambridge. But neither the

## What does the European Parliament do?

The European Parliament is a law-making body based mainly in Strasbourg, France, and made up of representatives (MEPs) who are elected for five-year terms by the 27 member states of the European Union. The number of MEPs from each member state varies according to the size of the country's population.

The parliament has the power to accept, reject or amend legislation, except in key areas such as taxation, agriculture and defence. Unlike national parliaments, it does not directly propose new legislation — that is the remit of the European Commission, based in Brussels, Belgium, which is divided into a number of departments.

The European Union Council of Ministers, made up of one representative from each of the member states, also meets in Brussels, and is the power broker of the triumvirate, responsible for approving all EU legislation. **Natasha Gilbert**

# Quaternary geologists win timescale vote

## Redefinition rescues once-threatened terminology from extinction.

In 2006, astronomers reached a decision on the planetary status of Pluto; now, geologists may have done the same for the status of the Quaternary, the time period in which humans evolved and live today. But, as was the case with Pluto, resolving this long-standing controversy has left some researchers feeling alienated.

The International Commission on Stratigraphy (ICS) has elected to formally define the base of the Quaternary at 2.6 million years before present, and also to lower the base of the Pleistocene — an epoch that encompasses the most recent glaciations — from its historical position at 1.8 million years to 2.6 million years ago. The decision, finalized on 21 May, will now be passed to the executive committee of the International Union of Geological Sciences (IUGS) for ratification, which is expected in the next month or two.

The vote shifts an 800,000-year slice, formerly part of the Pliocene epoch, into the Pleistocene. "It's kind of a land grab," says Philip Gibbard, a geologist at the University of Cambridge, UK, who has fought for the redefinition since 2001. "But we see it as just putting straight a mistake that was made 25–30 years ago."

In 1985, the beginning of the Pleistocene was defined at 1.8 million years ago, calibrated to an outcropping of marine strata in southern Italy. But some geologists have long felt that was a localized, arbitrary boundary that did not reflect worldwide changes — and argued instead for the 2.6-million-year mark, when the entire planet cooled.

The term Quaternary was adopted in the early 1800s, when geologists divvied up fossil records of Earth's history into four periods: the Primary, Secondary, Tertiary and Quaternary. The first two terms were discarded long ago, and although Tertiary is still sometimes used, in recent decades some geologists came to consider the Quaternary an outmoded relic. In 2004, a major publication left the Quaternary out of the ICS timescale altogether, making it vulnerable to extinction from scientific nomenclature. In place of the Quaternary, it extended the prior 'Neogene', which began 23 million years ago, up to the present. The Quaternary

community went into open revolt.

"The geologic timescale is fundamental for expressing the history of the Earth," says Stan Finney, a geologist at California State University in Long Beach and chair of the ICS. "This is our clock — we need the units of our timescale and their boundaries to be precisely defined."

Finney inherited the debate when he took his post at the ICS in 2008, and he vowed to come up with a democratic process to resolve it. After several months of open discussions and formal proposals from the Quaternary and Neogene communities, two rounds of voting took place, in April and May. The redefinition

proposal passed with approval from 16 of the 18 voting members.

Although for some the debate is settled, others are not pleased. "We don't take a metre stick in Paris and add a foot-and-a-half to it," says Lucy Edwards, a marine geologist with the US Geological Survey in Reston, Virginia. "You can redefine it by being more precise, but you don't increase its size by 40%." Edwards has practical concerns as well: in the 1980s, the USGS reworked all of its maps and terminology to reflect the decision to place the Pleistocene at 1.8 million years ago. Now that the international standards have changed, it will have to do so again.

Marie-Pierre Aubry of Rutgers University in Piscataway, New Jersey, who lobbied against the change, says that the rules of science are being violated. Whereas other major boundaries in Earth's history are associated with faunal extinctions and turnover, she says, "you come to the Neogene–Quaternary boundary, and there is nothing there". She notes that the term Neogene, not Quaternary, is used widely in textbooks to describe the current period. The Neogene community has already responded by petitioning the IUGS to suspend the vote.

Others are moving on. "In the end, it is only a shift in nomenclature," says Martin Van Kranendonk, a geologist at the Geological Survey of Western Australia in East Perth, and one of the two voting members who voted against the Quaternary proposal. "The rocks and time itself haven't changed," he says. "It's just what we have chosen to call them." ■

**Amanda Leigh Mascarelli**

## REDRAWING GEOLOGICAL TIME

The new definition of the Quaternary, as approved by the International Commission on Stratigraphy.

| | | |
|---|---|---|
| Present day | | Holocene |
| | Quaternary | Pleistocene |
| 1.8 my | | |
| 2.6 my | | |
| | Neogene | Pliocene |
| | | Miocene |
| 23 my | | (not to scale) |

J. BLAIR/CORBIS

ICS
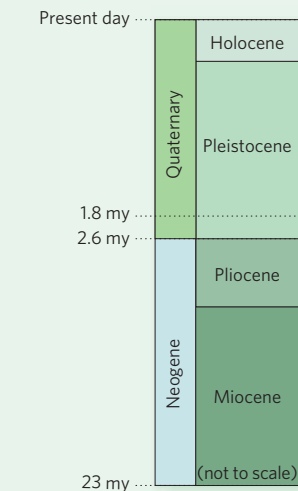
# Spooky research cuts

Two leading researchers in quantum computing have had their funds cut off by a US intelligence-research agency in what seems to be an administrative technicality. The controversy has underscored some scientists' fears that the field in the United States is too dependent on the spy world for funding.

In a letter dated 13 May that was sent to John Holdren, director of the White House Office of Science and Technology Policy, 12 researchers in quantum information science express their concerns about the US government's commitment to the field. The letter focuses on the sudden loss of support from the Intelligence Advanced Research Projects Activity (IARPA), which had previously funded two research teams at the National Institute of Standards and Technology (NIST). The groups are led by physicists David Wineland at NIST's facility in Boulder, Colorado, and William Phillips, a Nobel laureate whose lab is at NIST headquarters in Gaithersburg, Maryland.

Quantum computing has long been of interest to the US intelligence community because of its potential use for code-breaking. And intelligence funding bolstered quantum information science as it was getting going in the 1990s. But "the current approach of the intelligence community does not have a long-term view," says quantum information researcher Christopher Monroe of the University of Maryland in College Park, and it controls "a large fraction of the US effort in this field". The National Science Foundation, the largest single source of civilian government funding for the field, contributes US$15 million annually.

In 2007, the newly created IARPA took over funding for quantum information science from the National Security Agency. The following year, IARPA stopped funding the NIST researchers because, it says, it did not want to fund other government agencies.

NIST scientists declined to comment on the issue, deferring questions to IARPA. The intelligence agency confirmed that the funding had stopped, but declined to say how much money had been going to NIST, or the total amount of support for the field. Lisa Porter, IARPA's director, says it has been reviewing all of the projects it inherited. In

**"Having funding terminated based on a technicality seems incredibly shortsighted."**

the case of quantum information science, she says, the NIST funds ran out while managers were reviewing the programme and deciding how NIST might be involved. "Until that role is determined, three NIST researchers who were receiving funds from the legacy efforts are not receiving IARPA funds," Porter says. (A third NIST researcher, Ray Simmonds, lost funding along with Wineland and Phillips.)

Bureaucratic mix-up or not, some researchers argue that the decision weakens the field. "If funding for world-class research groups is altered so abruptly that postdocs must leave, graduate students be dropped, or planned experiments be cancelled, they would quickly deteriorate," says Daniel Kleppner, a physicist at the Massachusetts Institute of Technology in Cambridge, and one of the letter's signatories. "The recent concerns about the proposed funding changes from IARPA arose because there was not enough time to plan for alternative support, and world-class groups would have been seriously disrupted."

NIST spokesman Ben Stein says that the institute will continue to support the groups: "It's ultimately the responsibility of NIST to make sure these programmes receive the resources they need, and we are committed to ensuring they remain adequately funded."

And Porter argues that IARPA's support for the field is going up, not down, pointing to several new programmes that have been started since the agency was created. "In the aggregate, IARPA anticipates spending about 50–60% more in fiscal year 2009 than in fiscal year 2008 in the area of quantum information science," she says.

But others say the decision seems to run counter to a report on quantum information science released earlier this year by the National Science and Technology Council, which praised the work and called for a more coordinated strategy to maintain US leadership in the field.

"Anyone who hears about this is shocked beyond belief," says Ivan Deutsch of the University of New Mexico in Albuquerque. "The world leader in quantum computing having funding being terminated based on a technicality seems incredibly shortsighted." ∎

Sharon Weinberger

# Sweden likely winner for neutron source

Sweden is claiming victory in a long-running battle to host Europe's next-generation neutron-science facility.

The European Spallation Source (ESS), a €1.4-billion (US$2-billion) neutron facility, will be built in Sweden, according to officials there. The site, near the city of Lund, beat rival locations in Spain and Hungary.

Seven nations out of twelve potential backers voted for the Lund site. Only one, Portugal, voted for the Spanish location, according to Colin Carlile, the director of ESS Scandinavia, the organization in charge of the Swedish bid.

But the other candidate host nations are unwilling to concede defeat. The 28 May vote was simply an "opinion", says Lázló Rosta, the project director for ESS Hungary and a scientist at the Budapest Neutron Center. "There is sort of a Swedish advantage, but there is not a final decision," he says.

Spallation neutron sources, which accelerate protons onto a heavy metal target to produce neutrons, can create higher-powered neutron beams using less energy than conventional production methods using reactors.

The ESS has been on European neutron scientists' drawing board for the best part of a decade. Germany had originally offered to host the site, but later withdrew after the project was not endorsed by the nation's science council (see *Nature* **418,** 262; 2002). Since then, it has been in search of a home. In an unusual move, Spain and Hungary had promised to endorse each other's sites if either was chosen over Sweden.

But Sweden seems to have won. Carlile says that under the terms offered to the research ministers, Sweden, together with its Nordic and Baltic partners, will pay 50% of the project's construction costs. The team is looking for at least 35% of the funding to come from other nations, such as the United Kingdom, France and Germany. The remainder, up to 15%, will be paid for with a bridging loan from the Swedish government.

Rosta says that he believes there may be a way to include the Hungarian and Spanish proposals in the Swedish bid. The idea, he says, would be to develop the ESS accelerator in Spain and the instrumentation in Hungary. Alternatively, he says, one of the two sites could play host to a smaller spallation neutron source. "We would like to see the three sides arrive at a win-win situation," he says.

Carlile says that he hopes an agreement on cost sharing will come within months, so that construction can begin in 2012. That schedule would see the ESS delivering its first neutrons between 2018 and 2019. ∎

**Geoff Brumfiel**

> "We would like to see the three competing sides arrive at a win-win situation."

## University wins injunction against animal activists

Five animal-rights activists and associates were permanently barred last week by a California court from harassing researchers at the University of California, Los Angeles (UCLA).

The permanent injunction, awarded on 27 May, follows a temporary restraining order made in February last year (see *Nature* 451, 1041; 2008), which was converted into a preliminary injunction that April. It comes after repeated threats, including home invasions and fire bombings, aimed at UCLA researchers who use animals in experiments.

The activists were identified as associated with the Animal Liberation Front, the UCLA Primate Freedom Project or the Animal Liberation Brigade. Two of them face multiple felony counts for an alleged conspiracy of stalking and threatening UCLA researchers; their pre-trial hearing will be held on 14 July. Both pleaded not guilty.

## Economic gloom threatens renewables investment

Renewable-energy projects are struggling the most in terms of investment in the energy sector, according to a report published by the International Energy Agency (IEA) on 27 May.

The agency predicts that a sharp fall in energy demand could see global electricity consumption drop by up to 3.5% this year — the first contraction since 1945. Combined with tighter access to credit, the weak demand means that many firms in the oil, gas and coal sectors are having to cut back on their spending. The slump is even more pronounced for renewables projects, which are often developed by smaller companies.

Government stimulus spending will



SOURCE: IEA

MORE STIMULUS NEEDED

*IEA estimate

Legend:
- Wind
- Solar
- Biomass
- Marine & small hydro
- Geothermal
- Impact of stimulus packages*

trigger additional public and private investment in renewable energy (see chart), but the agency thinks that the stimulus funding needs to be increased sixfold if patterns of energy use are to be altered sufficiently to keep the global temperature rise within 2 °C of pre-industrial levels.
**See also page 740.**

## Malaria vaccine enters phase III clinical trials

GlaxoSmithKline's RTS,S malaria vaccine entered its final phase of pre-approval testing last week in Bagamoyo, Tanzania, raising hopes that the drug could be licensed for widespread use by 2012.

On 26 May, five infants aged 5–17 months were inoculated with the vaccine. In the coming months, the study will become the largest ever trial of a malaria vaccine candidate, involving 16,000 children under the age of two at eleven sites in seven African countries.

RTS,S is so far the only malaria vaccine to make it to phase III trials, in a development process that has taken more than two decades and cost over US$400 million (see *Nature* 451, 1042–1046; 2008).
**For a longer version of this story, see http://tinyurl.com/malvac.**

## Africa declares its stance for climate-change talks

More than 300 African negotiators, ministers, experts and agency representatives converged on Nairobi last week to forge a shared vision on climate change. Their conclusion: to ask for more financial and technical support to help Africa cope, but with few specifics.

"It was difficult to agree on how much to ask for," says Sputnik Ratner, a spokesman for Buyelwa Sonjica, South Africa's minister of water and environmental affairs, who chaired the ministerial meetings. "Different countries have different needs."

The 'Nairobi declaration' also asks industrialized countries to reduce their greenhouse-gas emissions towards the upper end of the pledged 25–40% cut below 1990 levels by 2020. The declaration is meant to serve as a starting point for Africa's negotiations at climate meetings being held this week in Bonn, Germany, and scheduled for December in Copenhagen, to hammer out a successor to the Kyoto Protocol.

Despite the declaration's vagueness, the United Nations Environment Programme hailed it as a "landmark position" and a "major milestone on the road for combating climate change on the continent".
**For a longer version of this story, see http://tinyurl.com/afclim.**

## Changes at the top for Indian science



I. MUKHERJEE/AFP PHOTO/GETTY IMAGES

**Prithviraj Chavan.**

India's Prime Minister Manmohan Singh last week named Prithviraj Chavan as his new science minister, after spring elections returned Singh's party to power (see *Nature* 459, 311; 2009).

Chavan, who has a master's degree in engineering from the University of California, Berkeley, takes charge of the ministries of science and technology and of Earth sciences. But he will continue as minister of state in the prime minister's office, a role he has held since 2004, and which involves interactions with the space and atomic-energy agencies.

"I am personally happy that we have a science minister who is also ideally positioned in the prime minister's office," says Thirumalachari Ramasami, secretary for the department of science and technology. With former science minister Kapil Sibal placed in charge of the human-resource development (education) ministry, Ramasami says the combination "signifies a very bright future for Indian science as a whole".

## Open-access publishing gains another convert

University College London (UCL) became the latest institution to adopt an open-access publishing policy this week, adding to the rapid increase in such mandates over the past year.

The policy, announced on 3 June, will see all of the university's research placed online in an institutional free-to-access repository — but only where publishers' copyright rules allow. The system will take effect from the beginning of the 2009–10 academic year.

"Open-access mandates [from institutions and funding bodies] have almost doubled globally in the year that has elapsed since Harvard's mandate in May 2008," says Stevan Harnad, an advocate of open access at the University of Southampton, UK.

---

**Correction**
In the News story 'Fusion dreams delayed' (*Nature* **459,** 488–489; 2009), the pie chart of ITER partners inadvertently omitted India. Like all non-European Union partners, India contributes 9% to the overall cost of the project. The html version of the story has been amended to show the correct chart, and a corrected PDF can be downloaded from http://tinyurl.com/nnwt9w.
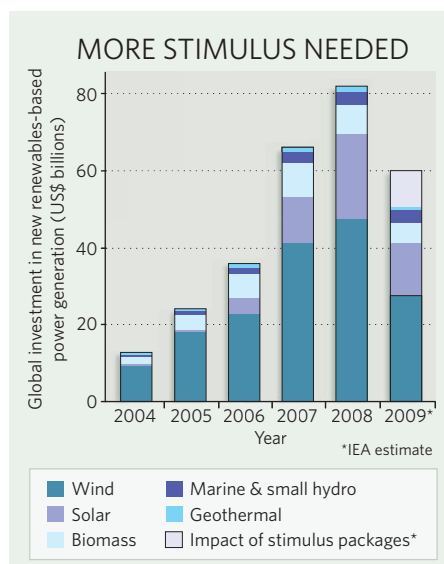
# Climate future

Deciding how to evaluate a cap-and-trade programme raises some thorny questions, says **David Goldston**.

Scientists sometimes describe the way human activities are changing the composition of Earth's atmosphere as a giant (and dangerous) experiment: we don't know enough about the complexities of the climate system to fully gauge all of the impacts of the increase in greenhouse-gas levels. Less frequently noted is that the antidote to global climate change is also a giant (although necessary) experiment: we don't know enough about the complexities of the economic system to fully gauge all of the impacts of a cap-and-trade system to limit greenhouse-gas emissions. These uncertainties are reason to pay attention to two little-noticed parts of the climate legislation now beginning to work its way through the US Congress: the section on research, and especially the section on programme evaluation.

Of course, the most noteworthy fact is that climate legislation is moving, although it is too soon to know if it will make it through the House of Representatives and the Senate. It was a watershed event when, on 21 May, the House Committee on Energy and Commerce approved a bill of almost 1,000 pages to cut US greenhouse-gas emissions (see *Nature* **459**, 493; 2009). This was the first ever vote in the House on a comprehensive climate bill and, more importantly, the first time Congress had worked on climate legislation with a president who would sign it into law. That's why industry groups, including those representing electric utilities and oil refineries, worked out deals with the bill's sponsors, Democrats Henry Waxman of California and Edward Markey of Massachusetts.

What industry wants most from a climate programme is regulatory certainty. Yet to be effective, a bill has to be open to alteration as more is learned about both the climate and the results of regulation. The Waxman–Markey bill includes several sections requiring evaluation of the cap-and-trade programme. Although this sounds simple, it raises some fundamental questions: how often should the programme be reviewed? Who should be examining it? What should they be evaluating? What should happen to their recommendations?

Most other climate bills assign the evaluation either to a government interagency committee or to the National Academy of Sciences. The advantages of a governmental panel are that its members may have first-hand knowledge of

**PARTY OF ONE**

how well emissions monitoring is working and what levels of emissions are being recorded, and that they can appropriately make policy recommendations. The advantages of the academy are its stature and presumed independence, but the assignment can take the academy into areas that strain its expertise and may require it to weigh in on questions that involve values as much as science. There is no one right answer, for example, when deciding how much money society should be willing to spend to fend off a possible consequence of climate change. Moreover, involving the academy in every aspect of the climate debate may erode its ability to act as an 'honest broker' on a range of issues — a vital function that few other institutions can fulfil.

The Waxman–Markey bill has it both ways, requiring the Environmental Protection Agency (EPA) to report to Congress every four years and then having the academy issue a review of that report the following year. The report would be extensive, including looking at US and international emission levels, their impacts and whether the emissions cap should be changed. This mix of report and review may be the best of both worlds, especially if the prospect of academy oversight makes the EPA more meticulous, but it could also set up an overt fight between the academy and an administration, or cause the academy panel to trim its sails to avoid one.

Under the bill, the academy would also provide an additional report every four years on "the emissions reduction potential, commercial viability, market penetration, investment trends and deployment" of technologies that could reduce emissions. Interestingly, the bill requires that academy panel to "include the

participation of technical experts from relevant private industry sectors". This is a reasonable idea and indeed perhaps the only way to tap the requisite expertise. But it also raises the spectre of conflict-of-interest charges and points to both how much of a stretch this report may be for the academy and how difficult it is for any report on the topic to be considered objective.

A year after the academy issues its two documents, the bill mandates that the president direct federal agencies "to take appropriate actions identified" in the EPA and academy reports. This verges on requiring the president to follow academy recommendations — a bad idea, given that policy recommendations involve more than science and should not be delegated to an unaccountable panel nominated by the academy. But the most substantial change that could be recommended — altering the overall emissions cap — would require congressional action.

A few scientists and science groups have weighed in on this language (and the committee staff asked me for my views early on without sharing any text), but this aspect of the bill merits fuller discussion — including whether new mechanisms need to be invented for reviews at the intersection of science, technology and policy. Climate regulation is not likely to remain fixed over the 40 years covered in the bill, and how possible changes are laid out for lawmakers may prove to be significant.

One facet of the bill that is likely to be the subject of greater debate in coming months is whether and how it funds research. Although President Barack Obama proposed setting aside US$15 billion a year for research and development, the Waxman–Markey bill includes only about $1 million a year to set up university research centres (although there is more money for technology deployment). Neither the Obama administration nor university lobbyists have pressed this issue yet, but university groups are planning to weigh in as the bill moves forwards. Funding in the Waxman–Markey bill is especially valuable because it would be available without further congressional action; it would not need to pass as part of the annual battles over spending legislation.

It's unsurprising that these provisions have been below the radar so far; the whole bill has received remarkably little front-page coverage. But research and programme evaluation need to be part of the discussion as the bill moves forwards — because no climate legislation can possibly get everything right the first time around. ∎

David Goldston (partyofonecolumn@gmail.com) is a visiting lecturer at Harvard University's Center for the Environment.

# MAGNIFYING POWER

New microscopes are revealing sights that have never been seen before. *Nature* profiles five machines that are changing how biologists view the world.

Close-ups of cork, lice and fly's eyes do not inspire the rhapsodies that they did more than 300 years ago when Robert Hooke first observed them under a microscope. But other pictures do — the boughs and twigs of a branching neuron in its forest; the scuttle of vesicles delivering molecular loads; the endless thrill of a cell carving itself in two again — and again, and again — as an embryo buds into being. Now, as then, microscopy is central to the understanding of living systems. In this special section, *Nature* reporters look at five microscopes that are resolving aspects of life in stunning new ways.

Microscopes today still do the job that Hooke asked of his: gathering information on details that the human eye cannot resolve and magnifying them to a size that it can.

Today's microscopes are used more and more to look at systems that are carefully prepared to make their workings visible. The dramatic fruits of this transformation can be seen in the green-fluorescent-protein revolution of the past decades, which has made it possible to engineer appropriate illumination into organisms under study. Its full impact will only be realized when microscopes can capture everything that these organisms can reveal.

This is why there is so much excitement around the arrival of the stimulated emission depletion (STED) microscope (page 638) and other 'super-resolution' techniques that are allowing light microscopy to resolve details on the nanometre scale, something once thought



The microscopes that greet undergraduate biologists in university teaching labs are remarkably similar to those that Carl Zeiss developed in the nineteenth century. On that basis of continuity, one might think that the technology was mature. But although the laws of optics have not changed, the ways they are applied are in constant flux.

Part of the flux is due to the ingenuity of engineers and scientists. Applying the laws of optics to electrons, rather than to light, was possibly the greatest development in microscopy of the twentieth century — and it is still yielding dividends in the twenty first, as the ultrahigh-voltage electron microscope at Osaka demonstrates (page 634). Recently some of that ingenuity has explored the possibility of doing with electronics what used to be done with carefully crafted glass, producing technologies that do away with lenses altogether. The microscope-on-a-chip featured here could turn microscopes into a disposable commodity (page 632).

But perhaps the greatest potential for progress is not so much in the engineering of microscopes, as in the engineering of what can be seen with them. Hooke and his successors used the microscope to see the world as it was, revealing seemingly miraculous detail at scales far too fine for human craft — powerful evidence, so it seemed, for the infinite superiority of divine craftsmanship.

impossible. And the unassuming single plane illumination microscope (SPIM) microscope, with its way of imaging life without killing it, could herald an era of 'systems microscopy' (page 630). New ways of manipulating life will supply — and indeed demand — new ways of seeing what is going on. At the same time, there will also be those who prefer to make observations without the interference of labels. For those researchers, there are improved ways to identify molecules by their intrinsic chemical properties, such as the stimulated Raman scattering microscope sitting in a Harvard basement (page 636).

All these developments share one thing in common: computers. As tools for the construction, manipulation and distribution of images, whether moving or still, in two-dimensions or three, computers are almost as central to the microscope now as the lens. The startled eye at the eyepiece, as rendered on our cover, may be increasingly a thing of the past, as all that microscopes show comes to be seen on screen. The shock of new discovery, though, will remain — and perhaps, even, intensify — for as long as the workings of life become ever more variously and acutely examined. ◼

**See Editorial, page 615, Essay, page 642 and online at http://tinyurl.com/microspecial.**
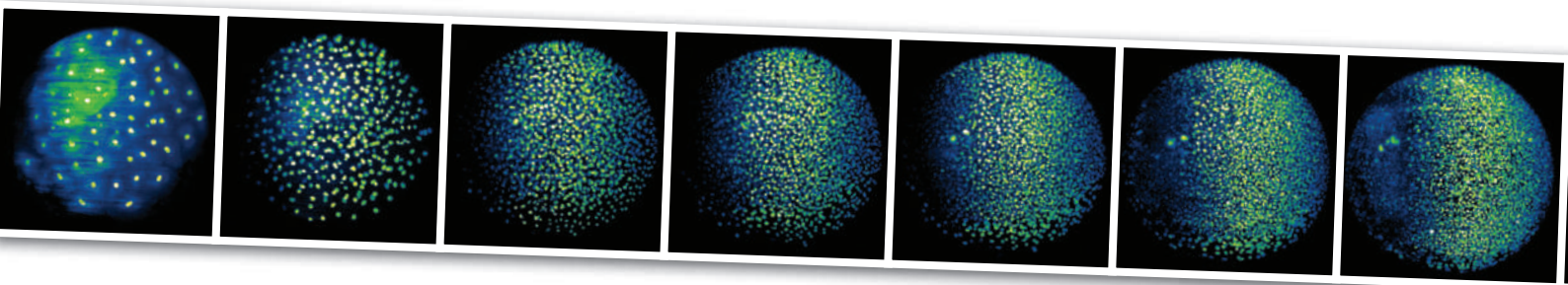
# SEEING THE SYSTEM
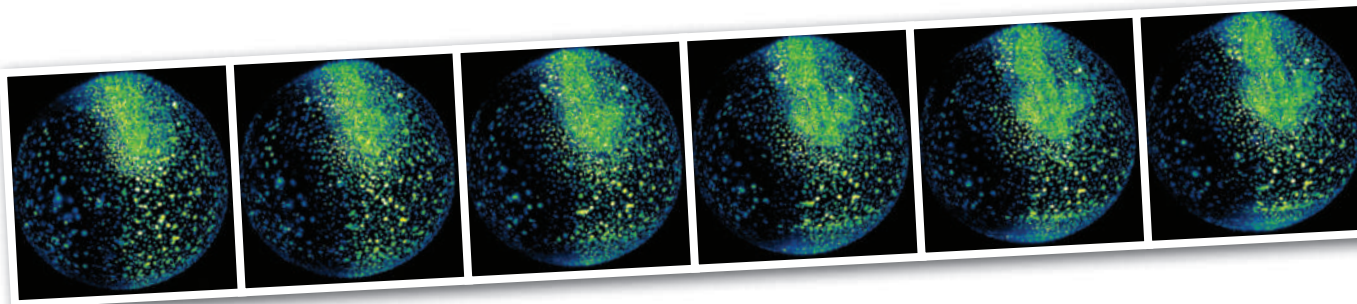
"We don't pretend to have invented new physics," says Ernst Stelzer modestly, standing next to an equally modest layout of lasers, mirrors and lenses. "It has all just been plain common sense."

Stelzer has been applying his common sense to microscopes ever since he arrived at the European Molecular Biology Laboratory in Heidelberg, Germany, in 1983 as a fresh-faced physics PhD student, and stayed on to head a team developing three-dimensional light microscopy. A quarter of a century in the field has now led to single plane illumination microscopy (SPIM), the technique with a modest face but an extravagant view: beautiful and unprecedented moving images of whole organisms as they grow one cell division at a time.

When Stelzer started out in the 1980s, a method called confocal fluorescence microscopy was beginning to show huge potential. The technique exploited the labelling of molecules in cells with a fluorescent tag to build up a three-dimensional image. But it still relied on traditional light microscopy procedures by which the biological sample is attached to a two-dimensional slide and a beam of light passes through the lens and through the entire sample, stimulating fluorescent emission that is detected by a camera.

Two problems with this approach bothered Stelzer. First, the squashed, two-dimensions of the slide were an unnatural environment for cells that live in three-dimensional tissue. And second, fluorescent microscopy slowly destroys the very thing that researchers are trying to make visible. The undiscriminating beam illuminates and triggers fluorescence from the entire sample, not just the focal plane. Yet photons are damaging to cells, which — apart from those

> **"This idea had been around for a century, but biologists didn't realise its potential."**
> — Ernst Stelzer

in skin, eyes or other body surfaces — receive little daylight. Light also 'bleaches' the fluorescent tags, which limits the number of times that the sample can be observed.

Stelzer determined to develop a microscope that could visualize living biological samples, for long time periods, in conditions that approximate normal physiology. To optimize the use of light, he turned a laser-light source at 90° to the camera and lens. In 2002, he added the element that is central to SPIM — mechanics that create a micrometre-thin sheet of light in a single plane that sweeps gradually through the specimen (see graphic). "This idea had been around for a century," he says, "but biologists didn't realize its potential for high-resolution." Only the fluorescent tags in this plane of light are excited, and the camera efficiently captures the photons emitted. The specimen is then rotated and the procedure repeated along half a dozen or so additional axes. Eventually all the planar images are merged together computationally into a three-dimensional whole. By eliminating excess illumination and dye photobleaching, the whole process can be repeated every two minutes or less for more than 24 hours. The specimen to be viewed — intact, in three dimensions — passes the time in a tiny, transparent cylinder filled with agarose gel to dampen any movement and perfused with physiological levels of gases such as oxygen and carbon dioxide.

Stelzer, working with colleague Joachim Wittbrodt, has shot spectacular movies of the first day in the life of a zebrafish embryo, a species loved by microscopists for its transparent skin (P. J. Keller *et al. Science* **322,** 1065–1069; 2009). As development is triggered, a single oval cell undulates, seeming to squeeze out additional complexity. More
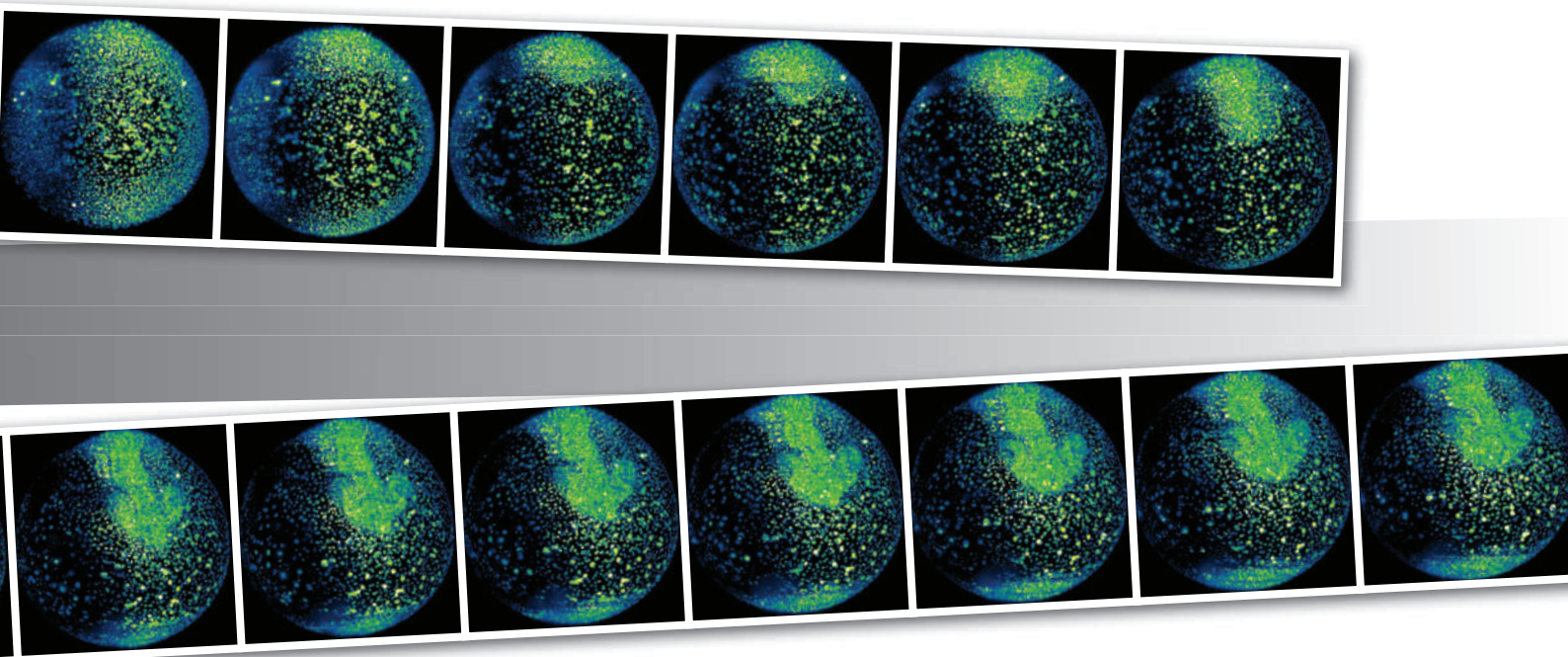
**Hourly snapshots taken during during zebrafish embryo development, starting 100 minutes after fertilization (top left).**

and more new cells shoot out along purposeful paths in a coordinated and symmetrical way. Emerging shadowy shapes soon become recognizable as organs. It all happens at a dazzlingly swift and confident pace until the one cell has multiplied to some 20,000 cells.

"It gives you a good feeling to see what is happening — you understand things more intuitively," says Carl-Philipp Heisenberg, from the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany. But Heisenberg says that the real wonder is in the data rather than in the movies. "That is where the real information is hidden — it is immensely helpful that the technique allows you to track all of the cells all of the time." For developmental biologists interested in tracking the destiny of each and every cell, these global, dynamic data are a boon, he says.

## SHEET LIGHTING
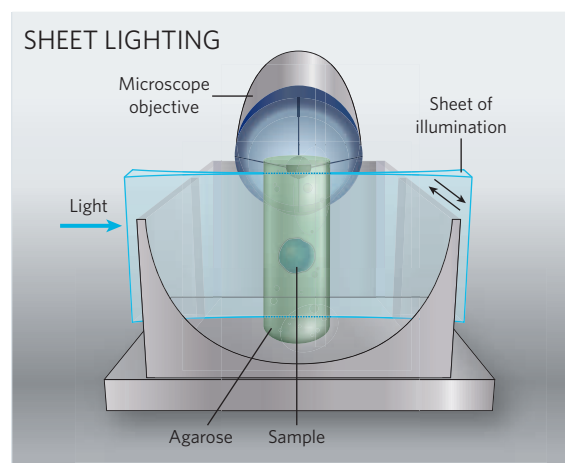


Stelzer's dream would take this further — into the realm of systems biology. He sees a future in which movies of thousands of developing embryos, each with one gene mutated, are recorded and compared. "Now we can think about systematizing the process to see the consequences of every possible gene mutation on development of the embryo, or of organs — or the effect of small changes in conditions such as temperature or acidity," says Stelzer. "It's a systems-biology microscope," agrees Wittbrodt, who talks about mapping the expression of genes on to the system too.

The process is computationally very demanding. Twenty-four hours of data for the zebrafish embryo took around 1,000 hours to process using their system. Stelzer is working on robust hardware and software so the microscope can be used by non-specialists. He and others are also refining the ability of SPIM to record movies of biological processes in real time. "Already you can see how an organ such as the beating heart develops, or how blood vessels develop, in the embryo", says developmental geneticist Didier Stainier, who is working with Stelzer's former PhD student Jan Huisken at the University of California, San Francisco.

Back in Heidelberg, Keller puts an adult fly into the container and flicks a switch. A fuzzy image of its eye, from a single-plane shot, appears on the computer screen. A few seconds later he flicks another switch to line up the images from hundreds of planes, and an intensely detailed view of the eye materializes like a rabbit from a magician's hat. But "it is not magic", says Stelzer. "It's simple technology." ■
**Alison Abbott**

See also page 629 and online at http://tinyurl.com/microspecial.

# MICROSCOPE FOR THE MASSES

Blurry specks in the eye seem an unlikely source of inspiration for a revolutionary microscope. But 'floaters' — tiny debris that floats inside the eyeball — led Changhuei Yang at the California Institute of Technology in Pasadena to devise a microscope so small, cheap and mass-producible that it could, according to its inventor, transform the way that microscopy is done.

The human eye registers floaters when bright light casts the shadow of debris directly onto the retina. On the tiny 'optofluidic' microscope that Yang and his colleagues invented, the sample casts a shadow directly on to an array of commercial light sensors as it floats along a microfluidic channel (X. Cui *et al. Proc. Natl Acad. Sci. USA* **105**, 10670–10675; 2008). The sensors feed the projection pattern to a computer, which constructs an image using relatively simple image-processing software. The device itself is assembled using semiconductor fabrication techniques and is smaller than an American dime. When mounted into a device with a USB port so that it can transfer information to a computer, it is still just 3 centimetres square.

Yang says that his microscopes, which could cost as little as US$10 apiece, could have the same revolutionary impact on science that the integrated circuit has had on the electronics industry. "When people were building transistors individually, it was still a pretty expensive proposition to build circuits out of them," he says. "But the move to build integrated circuits moved the semiconductor industry forwards because you could build things comparably cheaply with high functionality. If we can start to put 10–100 microscopes on a single chip and link a bunch of them up to operate in parallel to do high-throughput processing of a large number of samples, this opens up the opportunity to do experiments you might not otherwise do."

With cheap, high-throughput imaging, researchers could perform drug assays, genomic or proteomic screens and rapidly observe the outcome of hundreds or thousands of manipulations on the shape or behaviour of living cells. "It's very clever work," says Charles DiMarzio, director of the Optical Science Laboratory at Northeastern University in Boston, Massachusetts. "This is a way of making a [high-power] microscope that is very low cost, maybe even disposable, and that's something that we haven't had before."

Yang recognized that it would be difficult to shrink the lens and other delicate optics in a high-end instrument — so his 'direct projection' technique did away with lenses altogether. Other scientists have worked out similar techniques before, but they couldn't resolve anything smaller than 5 micrometres, because that's as small as the pixels on most digital light-sensing chips get. Yang coated the sensing chips with a thin layer of metal, than punched 500-nanometre holes into the metal to create apertures that are smaller than a pixel and that are patterned along the path of the microfluidic channel (see graphic). As the sample floats along, the chip captures repeated but staggered snapshots of what is passing overhead.



*X. CUI ET AL.*

With 500-nanometre holes, Yang's optofluidic microscope has a resolution that approaches that of a standard laboratory light microscope. He has already shown that it can capture images of the nematode worm *Caenorhabditis elegans* that are almost indistinguishable from those collected with a 20× objective lens on a conventional instrument. He is working to narrow the holes to 300 nanometres, a resolution capable of distinguishing the finer details of cells.

Besides transforming research microscopy, Yang's microscope could boost low-cost science and medicine in developing nations. The scope is rugged, works with sunlight, needs only the amount of computational power found in an iPod and, Yang wrote in his paper last July, might be "a boon for a health worker who needs to travel from village to village". That statement struck home for Ricardo Leitão, a postdoctoral fellow at New York University School of Medicine, who is founding a non-profit group called Tek4Dev — Science & Technology for Sustainable Development — which is putting together a tool kit to enable 'telemedicine' (using networks such as the Internet to facilitate clinical care) in poor countries. Leitão wrote to Yang on the day the paper was published to propose a collaboration. Yang, Leitão and Ana Rodriguez, a malaria researcher also at New York University, are now testing the microscope's ability to diagnose malaria-infected red blood cells based on their shape and those of the parasites inside them. Microscopy is still the standard method for diagnosing malaria, but microscopes can be few and far between in malaria-endemic areas. "Having a diagnostic tool as powerful as Yang's integrated with our hardware and 'tele' ability would be of tremendous clinical value," says Leitão.

Yang admits that by trying to do more with less, he is thinking differently from many of his peers. "In the field of biomicroscopy, there is a very strong drive towards building more sophisticated microscopes, giving you ever better resolution," he says. "But I think there's another axis to pursue, which is if you're actually building this in a comparably low-cost fashion, it can create experimental formats that are currently not doable using a traditional microscope or any other high-end microscope that other research groups are pursuing."

What is certain is that by making microscopes exceedingly small, Yang has actually been thinking very, very big. ∎
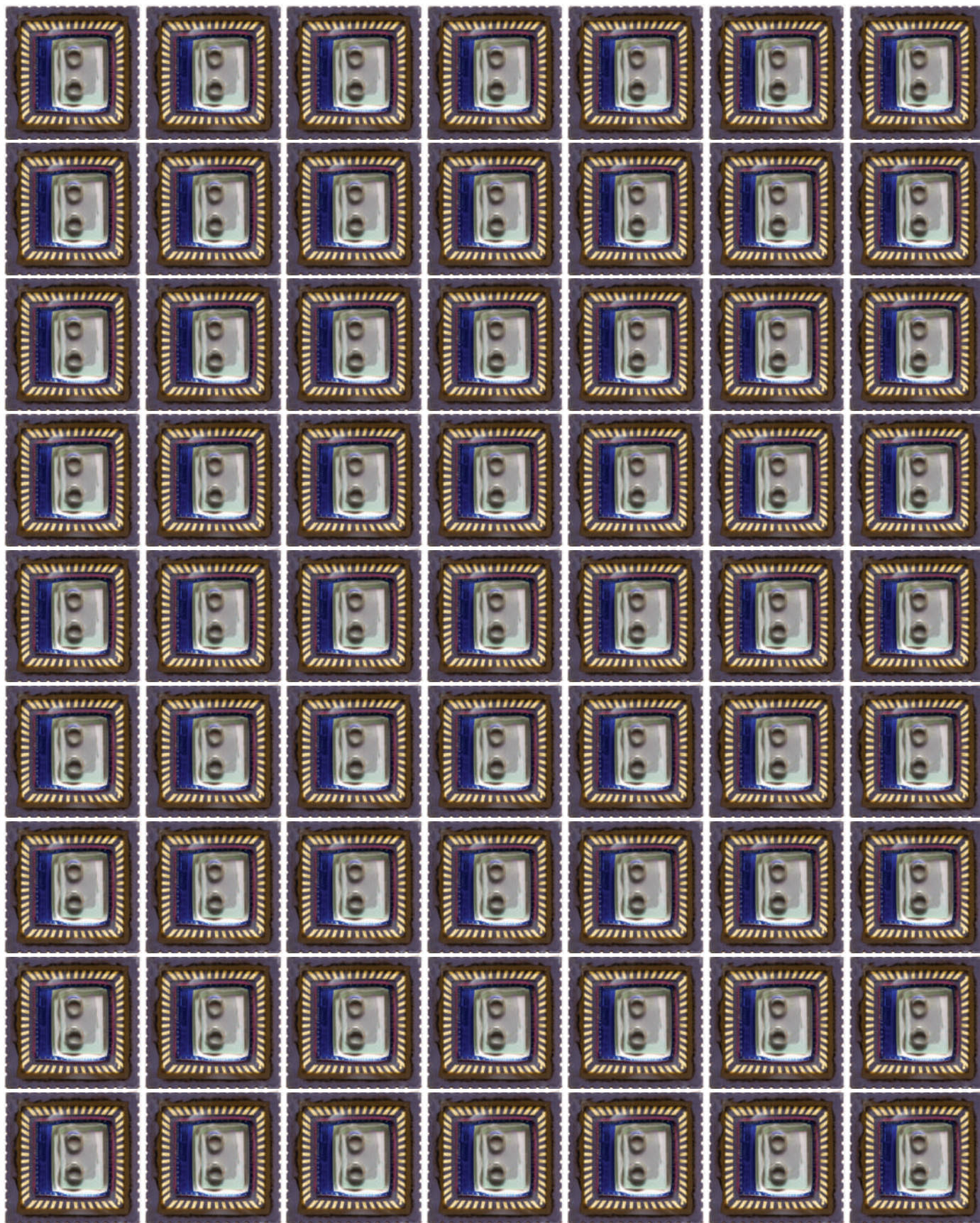
**Erika Check Hayden**
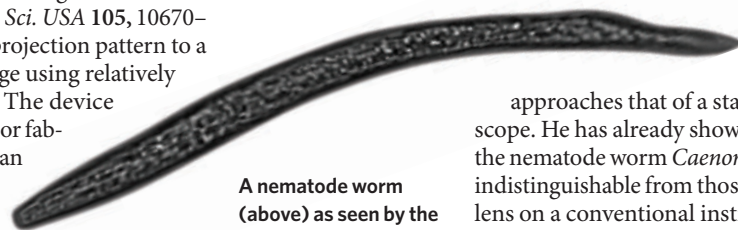
**A nematode worm (above) as seen by the optofluidic microscope (shown en masse and slightly larger than life, opposite, and in schematic below).**



*X. CUI ET AL.*

OPTOFLUIDIC MICROSCOPE

Specimen

Microfluidic channel

Apertures in metal layer
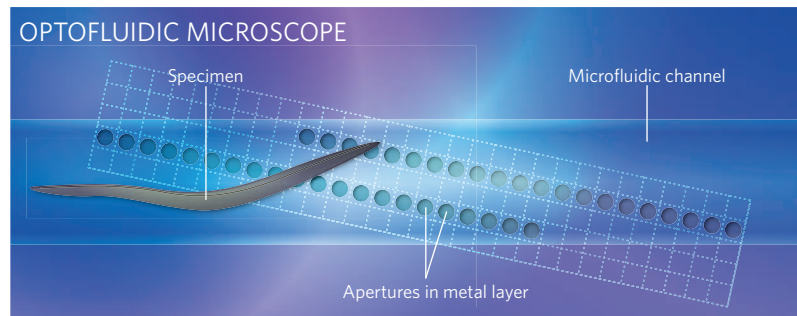
G. MARSHALL

Some say it looks like a rocket ready for take-off. Others say a pagoda. Either description seems more fitting for the tiered and towering 13-metre structure than the diminutive word 'microscope'.

The electron microscope at Osaka University in Japan is the most powerful of its kind. The technology behind it is nearly 80 years old, and the device itself has been in operation since 1995. But scientists there say that they are only just starting to show what this monstrous machine can do for biology.

Transmission electron microscopes shoot out a beam of electrons, which pass through some parts of a sample and are scattered by others. The emerging beam carries data that can be used to map out the structures it has transited.

into some 30 slices, making it difficult to ensure that a part was not being missed.

A decade ago, few life scientists used the Osaka UHVEM, says Hirotaro Mori, director of the Research Center for Ultra-High Voltage Electron Microscopy, which houses the machine. Materials scientists dominated them, especially those who were exploring the properties of thinly sliced semiconductors and integrated devices. But times have changed. Digital imaging has replaced film, slashing the time needed for data collection and analysis. An automated system makes it possible to better control the mechanics and flux of electrons and hence limit damage to tissues and cells. Computer software can correct much of the image distortion. And scientists have now developed electron-tomography

K. YOSHIDA / UHVEM, OSAKA UNIV.

# THE BIG AND THE BOLD

Because electron beams have a much shorter wavelength than visible light, they can easily resolve structures just a few nanometres across, far beyond the resolution of a conventional light microscope and better for viewing the details of cells' internal and external topography.

In the 1940s, scientists used some of the first electron microscopes to produce ground-breaking images of cells' mitochondria and other organelles. The detail was breathtaking, but researchers had to slice the tissue so that it was thin enough for the weakly energized electrons to get through: they were imaging dead, two-dimensional slices cut from living, three-dimensional structures. "Electron microscopy gave a framework to lay out physiological biology," says Mark Ellisman, a neuroscientist and microscopy specialist at the University of California, San Diego. "But it was kind of like taking road kill and analysing it."

Ultrahigh-voltage electron microscopes (UHVEMs), which give the electron beam enough energy to pass through and analyse thicker samples, started with two machines that could reach 3,000 kilovolts, built in Toulouse and Osaka in 1970. The new Osaka scope was constructed next door to that older one between 1991 and 1995, with a ¥2.4-billion (US$25-million) budget. It has an electron gun with a 5.7-metre accelerator tube and a maximum 3,500 kiloelectronvolt electron beam, and it can force electrons through biological samples thicker than 5 micrometres. This goes far beyond the 80-nanometre and 500-nanometre upper limits of the 100-kilovolt and 300-kilovolt machines typically found in laboratories.

The Osaka UHVEM is particularly valuable for imaging the diffuse but delicate structures of neurons, which requires an image that spans several micrometres at a resolution of tens of nanometres. Naoko Kajimura, an electron microscopist at Osaka University, has captured three-dimensional images of neuronal cells in the retina to show that those lacking a protein called pikachurin cannot form synapses (S. Sato *et al. Nature Neurosci.* **11**, 923–931; 2008). Had the 2-micrometre section been imaged by conventional electron microscopy, it would have needed to have been cut

> "Many people might say bigger is not better, but they don't have access to it."
> — Mark Ellisman

methods — in which the sample is tilted under the microscope's electron beam — which generate a series of images from different angles that can be compiled into an accurate three-dimensional structure. Now Mori says that biologists account for half of the machine's basic-research time, which has to be booked up to a month in advance.

Microscopists say that electron microscopy as a whole is undergoing something of a renaissance, with attempts to update and automate the technique so that it can image three-dimensional samples. The Osaka microscope is a brute-force solution. "Many people might say bigger is not better," says Ellisman, "but they don't have access to it."

Ellisman and his group, who operate the device remotely from San Diego, have been working to image a Purkinje cell, an expansive neuron that measures a couple of millimetres from the cell body to the tips of the exquisitely branched dendritic arbor. The group first took light microscope images of the cell, marking its location in the cerebral cortex so that it could be indexed in a cell-based brain atlas. The team then sent thick sections to Osaka to get nanometre resolution of the spines on the dendrites. Ninety per cent of the mammalian brain's synapses are made on these mushroom-like projections but, Ellisman says, "until now scientists have thought of them with tinker-toy-level abstraction". The UHVEM image, which shows the number of spines as well as the size and shape of the spines' heads and necks, will be crucial for simulations on how the brain works, he says.

Mori laments that such science experiments account for only half of the machine's use. To balance demand he gives the other 50% to industrial tasks, such as verification of newly designed integrated semiconductor devices, nanoscale characterization of new lubricants, or the measurement of hair damage after a cosmetic treatment. "I would like to see more basic science," says Mori.

**David Cyranoski**

M. ELLISMAN.
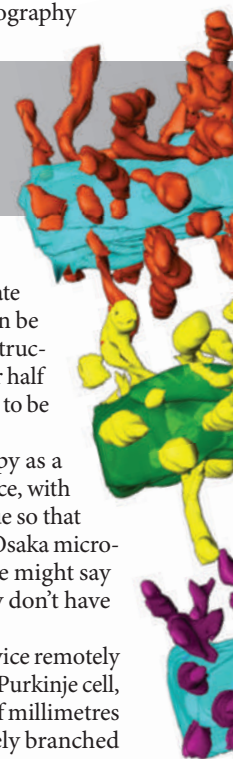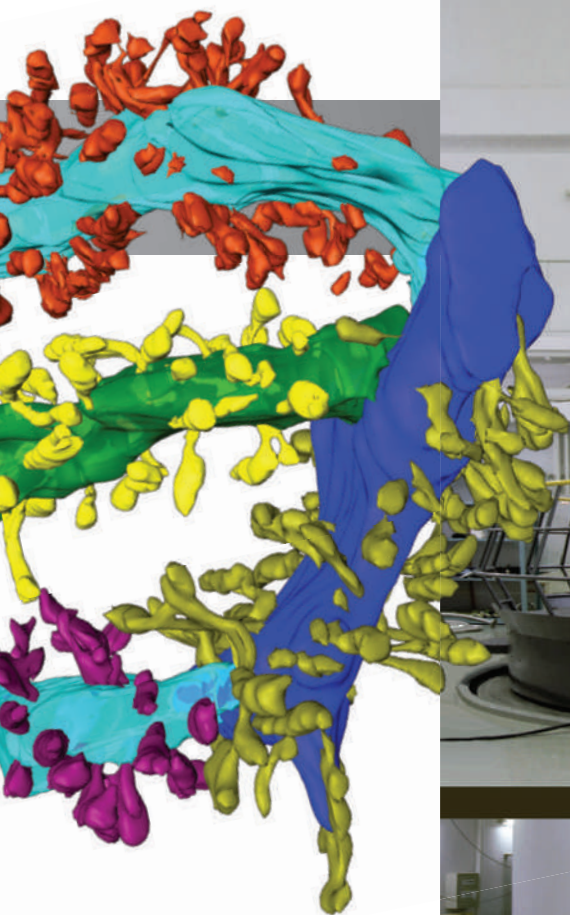
The fine-branching dendritic arbor of a Purkinje neuron (top) has been reconstructed from images taken by the Osaka electron microscope (right).

S unney Xie's newest microscopes don't look like the latest in sophistication. Tucked away in his biochemistry lab at Harvard University, they seem to be ad hoc assemblies of lasers, objectives and electronics, surrounded by a thicket of optical equipment. "Don't worry about most of these," says graduate student Christian Freudiger, gesturing to the latest addition to Xie's microscope family. "You only need a few optics to use the microscope. The rest are just for us to play with." As he leans briefly on the table, which is designed to eliminate vibrations, it counterbalances his weight with a reproachful "shhhh".

Freudiger and the other researchers 'playing' in Xie's lab have pioneered techniques to see biological molecules in their
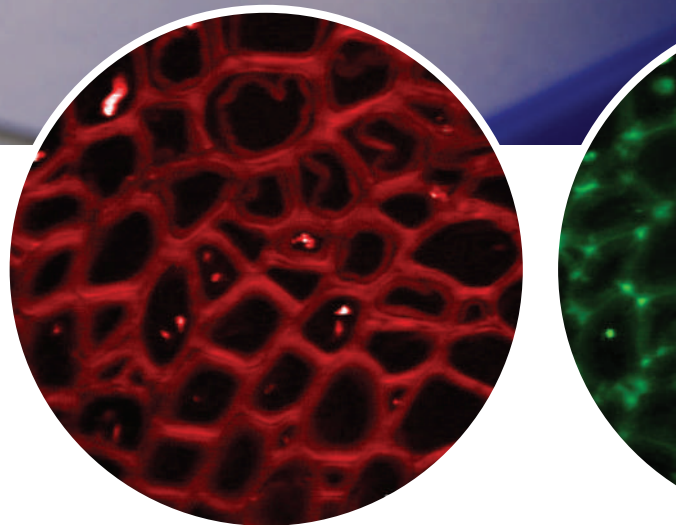
# THE NAKED MICROSCOPE

natural state. Based on Raman spectroscopy, the methods detect compounds from the characteristic vibrations of their chemical bonds, and they free the user from having to label molecules by attaching tags such as gold particles, antibodies or fluorescent proteins. Many researchers are unwilling to abandon labels because they offer an unparalleled ability to distinguish target molecules in a cell. But for others, who want to observe a molecule 'naked', without the interference of a molecular tag, it is proving to be hugely liberating.

In traditional Raman microscopy, laser beams illuminate a sample and the characteristic shift in wavelength caused by chemical bonds helps researchers pinpoint the identity and location of certain molecules. The method is therefore best suited for imaging molecules with distinct spectroscopic properties, such as lipids. But in early Raman microscopes the

signal was weak and the technique required long exposure times (sometimes more than a day), precluding the possibility of monitoring biological processes that happen in minutes, seconds or less. The microscopes also relied on powerful lasers, which fry delicate biological samples.

About a decade ago, Xie's lab developed a method called coherent anti-Stokes Raman scattering (CARS) microscopy, which uses two laser beams to excite molecular vibrations and generates a stronger signal. This technique cut down on exposure time and laser power, but was plagued by a high background signal.

Then in December last year, Freudiger together with postdoc Wei Min reported a further improvement[1]. Called stimulated Raman scattering microscopy (SRS), the method excites molecules with two laser beams that have been cali-

U. LOSTER

**"Biologists have never been able to get this kind of information before."**
— Shi-You Ding

molecule found in plant cell walls, without having to stain cells with techniques that affect the distribution of the two compounds. "Even if you want to tear down a building, you need to know what the structure is," says Ding, "but in this case there was just no way to see it."

Ding teamed up with Xie's graduate student Brian Saar to create movies of cellulose and lignin dynamics, allowing him to find chemical conditions that will break down lignin and leave the cellulose intact. "It's a very powerful tool," Ding says. "Biologists have never been able to get that kind of information before." Xie's group is also collaborating with researchers at the pharmaceutical company Pfizer in New York, who have already used the technique to track the acne-treatment retinoic acid as it is absorbed by skin. Scientists at Unilever based in the Netherlands and the United Kingdom want to use SRS to study the effects of cosmetics on skin and the distribution of fats and proteins in foods without having to attach bulky labels.

At the moment, an SRS microscope costs in the order of half a million dollars, but Xie hopes that technological





Sunney Xie (left), Christian Freudiger and their label-free microscope, which has resolved (from bottom left) cellulose; lignin; water (blue), protein (red) and oil (green) in a soya drink; and lipids in brain tumours.

brated so that the difference between the frequencies of the beams matches the vibrational frequency of the molecule to be imaged. The result: only the target molecules are excited and the troublesome background is squelched. By eliminating the high background signal, the SRS technique promises to extend label-free microscopy to molecules that couldn't previously be detected.

Biologists are already lining up to give Xie's microscopes a try, along with similar instruments being developed by other groups[2,3]. Biofuels researcher Shi-You Ding of the National Renewable Energy Laboratory in Golden, Colorado, wants to use SRS in his pursuit of techniques to degrade cellulose, a major component of plant cell walls, into smaller sugars that can be used for fuel. His research has been frustrated by the lack of a method to distinguish cellulose from lignin, another

improvements will reduce the cost and complexity of the set-up. Zeiss and Leica, two major microscope manufacturers, have just licensed the technology, and aficionados who have set up their own CARS systems can tweak their microscopes to accommodate SRS, he says.

SRS still has a few kinks to iron out, says Jing Kang, a professor of medicine at Harvard Medical School. It should enable researchers to distinguish between different types of unsaturated fatty acids at biological concentrations within the cell — discerning heart-friendly omega-3 fatty acids from the less-healthy omega-6 fatty acids, for example. But so far, Kang and his collaborators in the Xie lab have not been able to do so. "In theory it should be okay," Kang says, "but whether it's really going to make it — that's another story." ∎

**Heidi Ledford**

1. Freudiger, C. W. *et al. Science* **322,** 1857–1861 (2008).
2. Ozeki, Y., Dake, F., Kajiyama, S., Fukui, K. & Itoh, K. *Opt. Express* **17,** 3651–3658 (2009).
3. Nandakumar, P., Kovalev, A., & Volkmer, A. *New J. Physics* **11,** 033026 (2009).

B. G. SAAR, Y. ZENG, C. FREUDIGER, G. VAN DALEN, M. VAN RUIJVEN, C. XIAO, C. FREUDIGER, G. YOUNG, S. KESARI
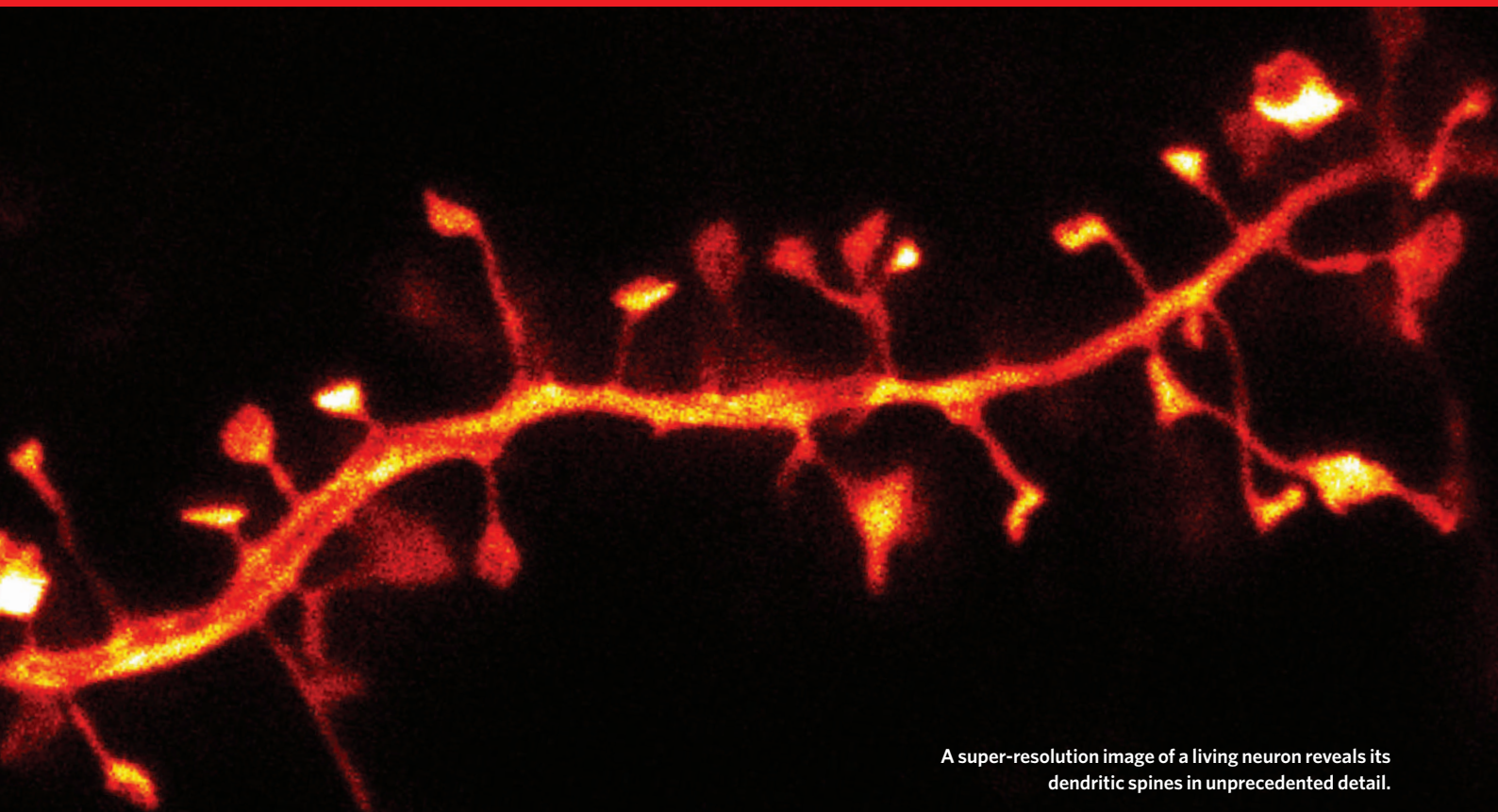
A super-resolution image of a living neuron reveals its dendritic spines in unprecedented detail.

" t's childish, but it still gives me great pleasure to see high-res pictures everyone told me would be impossible," says Stefan Hell of the razor-sharp silhouettes of mouse neurons on his screen.

Hell can't resist a small gloat. He frequently refers to his bitter struggle for recognition — the many years in the 1990s trying to reinvent microscopy as nanoscopy, able to reveal structures an order of magnitude smaller than light microscopy. 'You can't argue with the law of physics', was the no-can-do attitude of most microscopists at the time. They were referring back to 1873, when German physicist Ernst Abbe declared that diffraction inevitably limits the resolution of microscopy to around half the wavelength of light.

(see graphic). It hugs the excitation beam tightly, depleting all of the fluorescence in a sample except that at the very centre of each spot. This is the ingenious principle of stimulated emission depletion (STED) microscopy — and it throws objects measuring 20 nanometres into fine relief.

Hell dispelled much scepticism about his technique with a 2000 paper[1] in *Proceedings of the National Academy of Sciences* — "previously rejected by both *Nature* and *Science,*" he observes ruefully — which showed that his STED method could generate nanoscale fluorescent images. The objects themselves (beads, yeast and bacteria) were not exciting, but the proof of principle was. Two years later he was appointed a director at the Max

# THE GLORIOUS RESOLUTION

Hell didn't argue with the laws of physics — he found a way around them. Hell's innovation was to think about how he might exploit the properties of the fluorescent dyes, or fluorophores, that are widely used to label molecules for microscopy in biological samples. Fluorophores can exist in three states: ground, excited and dark. The confocal microscope, a standard machine in cell biology, works by rapidly scanning a sample with a focused beam of light and capturing the photons emitted when fluorophores are activated by the beam from ground to excited state. In practice, the 'diffraction barrier' means that two fluorescent spots less than around 200–300 nanometres apart appear as one blurred blob.

It dawned on Hell that he could use another beam of light, of slightly longer wavelength, to switch the fluorophores on the edge of the blurry spot back to their ground state. So he added a second beam to his confocal microscope. This beam is fashioned like a tube of light with a hollow centre
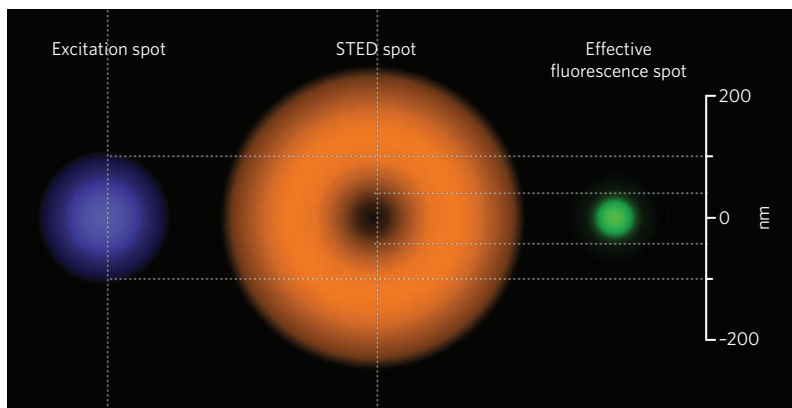


Superimposing a doughnut-shaped STED beam on the blue excitation beam shrinks the fluorescent area in a sample (green) to tens of nanometres, below the diffraction limit.

Planck Institute for Biophysical Chemistry in Göttingen, Germany, with all the research money and facilities he could handle. "There is still refinement needed — we are working on being able to do the same things at lower light levels and with a larger field of view," says Hell. "But the basic principle has been cracked."

Hell now finds himself with stiff competition as other labs race to develop, improve and commercialize techniques for 'super-resolution' microscopy. US scientists have developed methods known as stochastic optical reconstruction microscopy (STORM) and photoactivated localization microscopy (PALM), which use computation to work out the location of individual fluorophores and compile them into high-resolution images. And in structured illumination microscopy (SIM), biological samples are illuminated with stripes of light rather than a focused beam, and the interference patterns of the stripes are analysed and used to reconstruct images on nanometre scales.

Biologists can't team up with the new nanoscopy labs fast enough. "It is coming at a really exciting moment in biology," says Jeffrey Lichtman, a cell biologist at Harvard University, "and particularly in neuroscience, because neural connections are — tantalizingly — just below the diffraction limit." In his energetic bid to map all the neural connections in the brain, Lichtman says he is collaborating with all the main super-resolution microscopy labs.

STED microscopy has already revealed cell biological details that had been lost in the fluorescent fog. A video accompanying Hell's 2008 *Science* paper[2] shows 40-nanometre-diameter vesicles moving within a nerve cell. Nothing like this had ever been seen before. The speed with which the vesicles zap around, like fleas on a blanket, is almost unseemly. It is one thing to appreciate the theory and mathematics of cell dynamics, but it is



STED reveals the folds of the mitochondrial inner membrane, previously visible only with the electron microscope.

**The technology will be crucial for a large number of biological questions."**
— Tobias Bonhoeffer

us real biological insight — you can't achieve a dynamic view at this scale any other way," says Simons.

Although impressed by the technical advance, some researchers say that super-resolution microscopy has yet to prove itself. "I think these techniques owe us a real scientific breakthrough in solving a question that couldn't be solved any other way," says Winfried Denk, a microscopy pioneer based in at the Max Planck Institute for Medical Research in Heidelberg, Germany.

Others have no doubt that it will. For Tobias Bonhoeffer, a director of the Max Planck Institute of Neurobiology in Martinsried, Germany, STED microscopy is "a transforming technology which will be crucial for a large number of biological questions". In collaboration with Hell, he has published pictures of the tiny dendritic spines on neurons, which are central to processes such as information storage
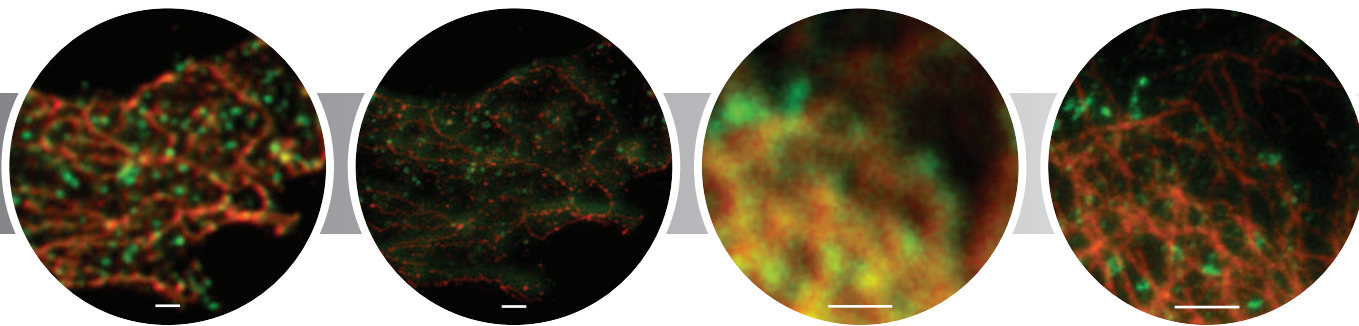


quite another to finally be able to see it in action.

Kai Simons, a cell biologist at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany, has seen his own 'lipid raft' theory gain a slug of support from the Hell lab[3]. Simons's originally controversial theory, proposed in 1988, held that certain lipids and proteins in a cell membrane form temporary clumps to, for example, facilitate the transmission of signals across the membrane[4]. But the proposed rafts would have been as small as 30 nanometres in diameter — invisible until STED microscopy came along. Hell's group showed certain proteins and lipids being transiently trapped in the complexes, exactly according to the theory. "It has given

**Side-by-side comparisons of human glioblastoma using a confocal microscope (first and third from left) and STED (second and fourth). Scale bars all 1 micrometre.**

in the brain, learning and memory[5]. "Spines are around a micrometre across — you can see them with light microscopy, but not the details," he says. "With this increase in resolution, we can even see the fine filopodia reaching out to make connections." ∎

**Alison Abbott**

1. Klar, T. A., Jakobs, S., Dyba, M., Egner, A. & Hell, S. W. *Proc. Natl Acad. Sci. USA* **97,** 8206–8210 (2000).
2. Westphal, V. *et al. Science* **320,** 246–249 (2008).
3. Eggeling, C. *et al. Nature* **457,** 1159–1162 (2009).
4. Nägerl, U. V., Willig, K. I., Hein, B., Hell, S. W. & Bonhoeffer, T. *Proc. Natl Acad. Sci. USA* **105,** 18982–18987 (2008).
5. Simons, K. & van Meer, G. *Biochemistry* **27,** 6197–6202 (1988).

# CORRESPONDENCE

## Fair-use policies aim to balance access and cost of publishing

SIR — Your News story 'Open-access policy flourishes at NIH' (*Nature* **458,** 690–691; 2009) raises the related question for publishers about the challenges inherent in providing the widest and most cost-effective access to quality scientific literature.

At the non-profit American Institute of Physics, we (like other publishers) are concerned about the inescapable realities of accomplishing this aim. The reality is that publishing a quality journal costs money. The trade-off is that those costs add value to a researcher's manuscript as it evolves from a draft submission to a final publication.

In our case, this transformation process requires peer review by tens of thousands of experts annually, as well as editing by postdoctoral physicists, refining of text and graphics, and use of increasingly sophisticated indexing and archiving measures. Government-mandated open-access policies will impose unintended negative consequences if they threaten the very business models that pay for quality publications.

Such mandates may not be necessary. We and other publishers are voluntarily developing copyright-friendly, fair-use policies that obviate the access issue. Harvard and the American Physical Society, for example, recently agreed on ways to facilitate authors' compliance with Harvard's new open-access policies when publishing in distinguished journals such as *Physical Review*, *Physical Review Letters* and *Reviews of Modern Physics*.

Even in the information age, maintaining the quality of the scientific literature costs money. How we can continue to improve access to scientific journals without compromising their quality is a question about which we should all be concerned.
**H. Frederick Dylla** American Institute of Physics, One Physics Ellipse, College Park, Maryland 20740-3843, USA
e-mail: dylla@aip.org

*Nature Publishing Group (NPG) provides a free manuscript deposition service at http://tinyurl.com/mq4mx2, to enable authors to comply with funders' mandates. NPG's licence-to-publish policy is at http://tinyurl.com/4etxvz*

## Funding ban could break careers at the toss of a coin

SIR — In your News story 'UK scientists get funding ban reprieve' (*Nature* **459,** 20; 2009), you report the response of the UK Engineering and Physical Sciences Research Council (EPSRC) to criticism of proposed changes to grant-submission eligibility. In my view, the EPSRC response still fails to address the central issue.

Even with the softened proposal, blacklisting will severely damage researchers' careers; moreover, it will selectively damage the productivity of our most innovative and daring researchers. It is critical that any measure used to do this should be robust, objective, transparent and widely trusted. Peer review as a means of numerically ranking grant proposals satisfies none of these criteria. Studies on peer review of grants (for example, S. Cole *et al. Science* **214,** 881–886; 1981) and of papers (for example, L. Bornmann and H.-D. Daniel *Learn. Publ.* **22,** 117–125; 2009) have consistently shown that peer review performs poorly at numerical ranking, that it is subject to serious random effects, and that it is not good at distinguishing between the majority of proposals and papers that fall between the very top and the very bottom. Correlation between results when a peer-review process is repeated is only marginally better than would be expected by chance.

Therefore, whether a proposal falls within the top or bottom 50% of the ranking is largely determined by chance, and has little relation to any objective measure of quality. The chief executive of the EPSRC has been reported as accepting that peer review is "basically a lottery" — a widely held view among researchers. However, the 50% mark is the key measure that the EPSRC proposes to use. The revised proposals are therefore still capable of potentially destroying people's careers on the basis of a measure that is only slightly better than flipping a coin.
**Cameron Neylon** Science and Technology Facilities Council, Rutherford Appleton Laboratory, Didcot OX11 0QX, UK, and School of Chemistry, University of Southampton, Southampton SO17 1BJ, UK
e-mail: cameron.neylon@stfc.ac.uk

## Cancer screening for women in developing countries

SIR — Your Editorial 'Early warnings' (*Nature* **458,** 679; 2009) points out some pitfalls in the effectiveness of cancer screening. I would like to draw attention to problems associated with gynaecological screening in resource-poor countries.

Disease prevalence can cause a large variation in the positive predictive value (PPV) of tests whose sensitivity and specificity are comparable. If 10% of the population is likely to have a disease and 1,000 people are screened for it, then by using a test with 100% sensitivity and 99% specificity, we should pick up 100 true positives and 10 false positives; the PPV will be 100 true positives to 110 total positives, or 91%. But for a disease with an incidence of 0.1%, a test of the same sensitivity and specificity will generate one true and 10 false positives per 1,000 screened, with a resultant PPV of one true positive/11 total positives, or 9%.

Take, for example, screening tests for ovarian and cervical cancers, to identify precancerous conditions and early-stage disease. Although these cancers are widespread killers, they have a low incidence compared with, say, tuberculosis. Ovarian cancer has a worldwide incidence of 0.01% (International Agency for Research on Cancer, www-dep.iarc.fr), meaning that one screen, even with 99.9% specificity, will give a PPV of 9%. The resulting surgery will be unnecessary for 10 out of 11 women who undergo it.

Such high specificity is very difficult to achieve, but anything lower is useless. Even multivariate systems combining many markers for ovarian cancer have failed to increase specificity above 99.9%. Despite this wastefulness, sponsors and governments are under pressure to organize large-scale trials, as your Editorial mentions.

The value of screening is also affected by the feasibility of further investigations and treatment. Take cervical cancer, which overall is more than twice as common as ovarian cancer (although its incidence varies widely around the world): the position of the cervix means that even a low PPV will not undermine well-established screening programmes, as results can easily be verified by biopsy rather than a major operation.

However, in many resource-poor countries, where only about 1% of women may be screened for these cancers, the confounding false negatives and negative predictive values caused by low sensitivity present a big problem. Many cases will not be discovered until the disease is too far advanced for successful treatment. In such countries, huge populations and restricted resources mean that screening efforts may need to be abandoned in favour of inexpensive universal treatment or prophylaxis. Newer, targeted approaches for ovarian cancer (T. A. Yap *et al. Nature Rev. Cancer* **9,** 167–181; 2009) and a cheaper vaccine for cervical cancer may prove helpful.
**Chinmoy K. Bose** Gynaecological Oncology Section, Netaji Subhash Chandra Bose Cancer Research Institute, Kolkata 700 006, India
e-mail: ckbose@hotmail.com

# ESSAY

# A microscopic reality tale

The earliest microscopes shed light on a once-invisible world. But, **Patricia Fara** explains, microscopists were uncertain about how well the images reflected reality — just as they are today.

In January 1665, Samuel Pepys followed up an afternoon with a prostitute by dining with the President of the Royal Society. On his way home, he called in at his bookseller and ordered an advance copy of Robert Hooke's book of the microscope. A couple of weeks later, Pepys collected his custom-bound volume and stayed up half the night absorbed in *Micrographia* — "the most ingenious book that ever I read in my life", he noted in his diary.

Pepys already owned a microscope, but he had never encountered images such as these. Like everybody who saw them, he was stunned by the intricate drawings of Hooke, who had modified a shop-bought instrument to reveal the minute details of insects and minerals, plants and feathers. For the first time, seventeenth-century gentlemen could closely inspect tiny yet only too familiar aspects of their daily lives — fleas, cheese mould, crystals of frozen urine, lice, nettle stings. In words whose punning eloquence complements his images, Hooke pointed out to his readers that a louse is "so proud and aspiring withal, that it fears not to trample on the best, and affects nothing so much as a Crown".

Hooke had originally presented his blown-up versions of formerly invisible objects in weekly meetings held at the recently founded Royal Society in London. The other Fellows there had encouraged him to publish his drawings in a book to advertise their new way of studying the world.

They had adopted as their figurehead the late Francis Bacon, who had been squeezed out of his post as Lord Chancellor in 1621 on charges of corruption. With his political career ended, Bacon had been free to dedicate himself to philosophy. He had set out to reject Aristotle's theoretical approach and insisted that the best way to learn about God's world was through observation and experiment.

In his writings, Bacon had stressed the Bible's message that humans are fallen creatures with defective vision; their comprehension is, he wrote, clouded by imagination. Attempting to overcome people's intrinsic weaknesses, Hooke invented devices for augmenting the senses — hearing and smell as well as sight. Indeed, *Micrographia* is an extended argument demonstrating that the unity and beauty of God's creation are best examined through detached, objective observation.

Hooke claimed that when experimenters looked through a microscope, all they needed to transcribe nature directly onto paper was "a

"For decades, biologists put themselves, not magnified specimens, in front of the lens."

*sincere Hand*, and a *faithful Eye*, to examine, and to record, the things themselves as they appear". To reinforce Bacon's biblical message of human fallibility, his first plate exposed the imperfections in objects produced by humans: under his microscope, a honed razor looked jagged, a needle's point became blunted, and a full-stop lost its sharp edge.

### Seeing is believing

Yet in practice Hooke found that "a *sincere Hand*, and a *faithful Eye*" were not enough to capture Nature exactly as she is. For one thing, microscope images were often ambiguous, so attempts to decipher them were shaped by individual expertise and personal interpretation. "The Eyes of a Fly in one kind of light appear almost like a Lattice, drill'd through with abundance of small holes," Hooke remarked; "In the Sunshine they look like a

WELLCOME LIBRARY, LONDON;
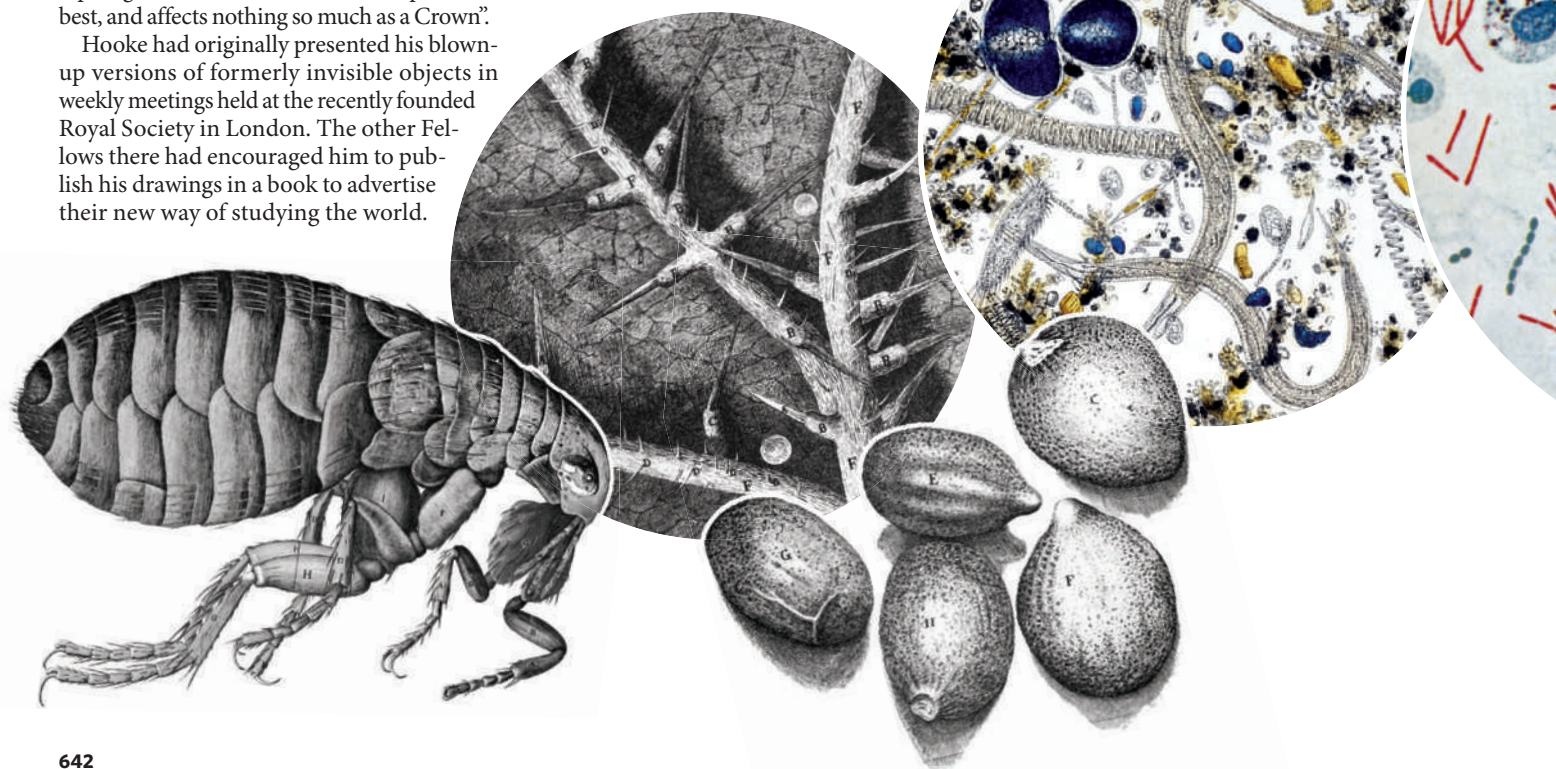
Surface cover'd with golden Nails; in another posture, like a Surface cover'd with Pyramids; in another with Cones." An observer needed a fine instrument, but also the skill to judge when it was yielding the right answers.

Hooke realized that further problems were introduced when he tried to reproduce what he saw on paper. *Micrographia*'s readers were removed from the original specimens by two intermediaries: Hooke himself and his engravers, whom he tetchily accused of making inaccurate copies. And unlike the colourful microscopic world that he observed through his lens, Hooke's illustrations were in black and white, mostly free-floating on the page rather than confined within the circular boundary of an eyepiece.

It is easy to assume that with better technology, Hooke and those who read his book might have come closer to seeing things as they truly are. But no straight line of progress links him with modern microscopists.

Hooke's eighteenth-century successors were so concerned with divine perfection that they often portrayed not the natural specimens they saw in front of them, but imagined versions that improved on observed reality. For instance, a particular ant under examination might have a missing leg or a deformed head, but illustrators would mentally combine all the ants they had ever seen to depict an ideal ant, as though they were trying to capture the essence of antness. Their approach was similar to that of anatomists of the same period, whose drawings of female skeletons were given particularly wide pelvises (thought necessary for child-bearing), while those of males were given especially large skulls.

Around the middle of the nineteenth century, biological fashion swung the other way: the rallying call now became 'Let Nature speak for herself.' Scientists began to insist that a faithful record should be made of the individual specimen under observation, including all its defects. By developing new types of microscope with vastly enhanced accuracy, they hoped to eliminate the subjectivity entailed in personal judgements.

> **"Even for scientists, clarity seems to trump neutrality."**

But like Hooke two centuries earlier, experimenters found it impossible to eliminate human involvement. On the contrary, as their instruments became more complicated, they had to intervene more. Under a stronger lens, less of the object could be seen at one time, which meant that more expertise was needed to identify the image and be confident that it represented external reality. Also, specimens now had to be carefully prepared, mounted and stained. Just as Hooke had to consider how his fly eyes were illuminated, so later microscopists had to ensure that light or dark spots were not instrumental artefacts.

### Article of faith

Photography, which was introduced in the early 1800s, was initially hailed as a direct recording technique that would eliminate human error and subjective interpretation — a veritable *Pencil of Nature*, as photographer William Henry Fox Talbot titled his self-promotional book (published in instalments
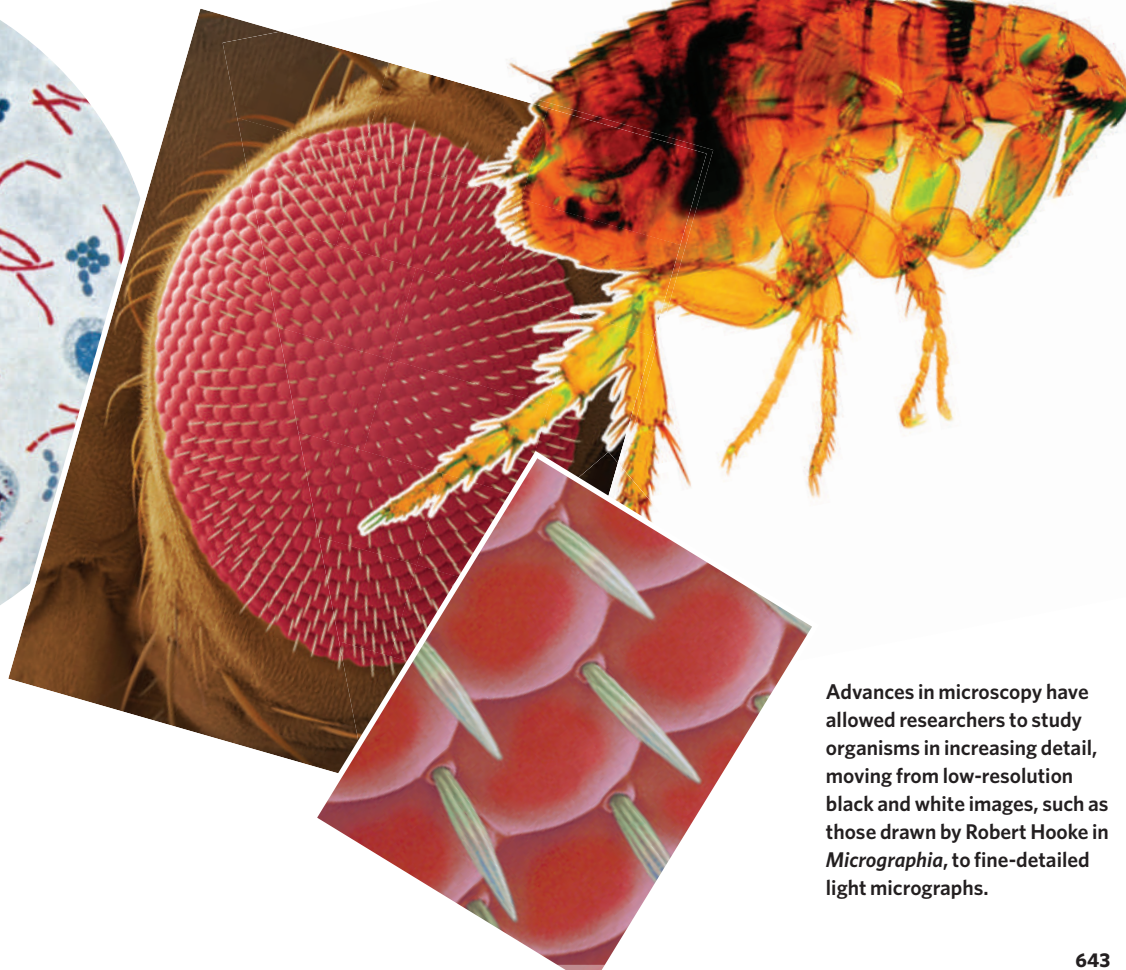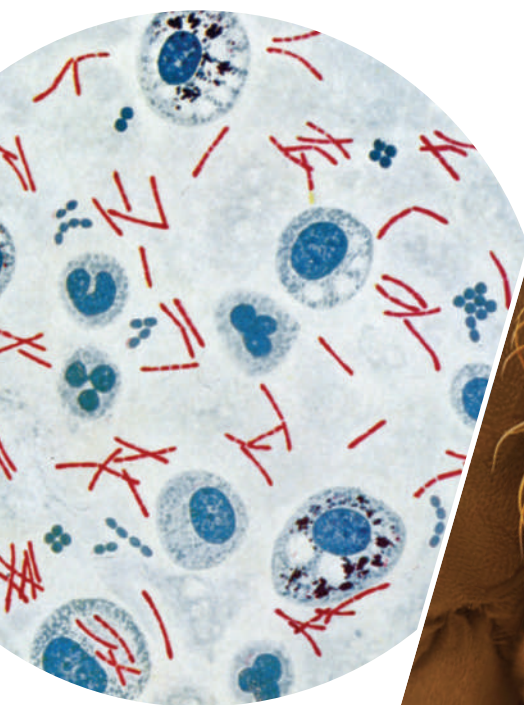
**Advances in microscopy have allowed researchers to study organisms in increasing detail, moving from low-resolution black and white images, such as those drawn by Robert Hooke in *Micrographia*, to fine-detailed light micrographs.**

**Microscopes have got more sophisticated, but have they brought us closer to seeing Nature as she really is?**

between 1844 and 1846). "No human hand has hitherto traced such lines as these," enthused the physicist Michael Faraday, an early campaigner for the use of photography in science; "What man may hereafter do, now that dame Nature has become his drawing mistress, it is impossible to predict."

But all this excitement didn't bear much fruit for microscopy, at least initially. For decades, biologists kept cameras out of their laboratories, and put themselves, not magnified specimens, in front of the lens by posing for publicity shots in artists' studios. In retrospect, this turn to self-portraiture may seem surprising, but scientists were keen to establish their public standing.

Even when cameras did begin to be used in laboratories, bitter arguments ensued among researchers about the pros and cons of using cameras as objective recorders. Many scientists felt that if people could believe they saw spirits returning from the dead in photographs, then the technology could hardly be relied on to reveal the truth of the microscopic world. Long exposure times made it difficult to capture things that moved, and uncertainties were introduced by lens imperfections or fluid splashes and the lack of colour. Moreover, it was impossible to make accurate measurements of the specimens photographed because development processes often shrank or stretched the photographic paper.

Printing limitations also meant that photographs had to be hand-copied as engravings before they could be mass-published, thereby allowing human involvement to creep back in. Engravers could highlight crucial features that were hard to pick out in blurred photographs packed with detail — but when does clarifying an image become distortion?

By the beginning of the twentieth century, mechanical reproduction of the images seen through the eyepieces of microscopes had won favour over drawings. "If a microscopist really did need a drawing," wrote the German researcher Erwin Christeller in his 1927 histological atlas of disease, then he should assign the task to an untrained technician, the closest approximation to a mechanical device. Mechanization spelt progress, and microscopes became increasingly sophisticated, revealing aspects of the natural world that had previously belonged to the domain of conjecture.

Hooke would have been thrilled by the three-dimensional images generated by today's electron microscopes showing the minutest striations on insect carapaces, yet his goal of eliminating human intervention has remained elusive. Don Fawcett, a pioneer in electron microscopy, stressed that technical progress depends on an inherent preference for order that cannot be factually demonstrated. It is "an article of faith", he wrote in an essay in the mid 1960s, "that an image which is sharp, coherent, orderly, fine textured and generally aesthetically pleasing is more likely to be true than one which is coarse, disorderly and indistinct". Even for scientists, clarity seems to trump neutrality.

As non-optical microscopes probe ever deeper into the once invisible world, viewers are further and further removed from reality. Nowadays, they are separated from the object itself by complex processes that transform electrical signals into visual representations made with artificial colours. If Bacon were confronted by the electronic image of a virus, he might well object that science suffers from what he called "Idols of the Tribe" — those errors of perception arising from a limited human mind that are "like an enchanted glass". ∎

**Patricia Fara** is senior tutor of Clare College, University of Cambridge CB2 1TL, UK. Her most recent book is *Science: A Four Thousand Year History* (Oxford University Press, 2009).
e-mail: pf10006@cam.ac.uk

FURTHER READING
Hooke, R. *Micrographia, or Some Physiological Descriptions of Minute Bodies* (Cosimo Classics, 2007 reduced facsimile of 1665 edition).
Bennett, J., Cooper, M., Hunter, M. & Jardine, L. *London's Leonardo: The Life and Work of Robert Hooke* (Oxford Univ. Press, 2003).
Tucker, J. *Nature Exposed: Photography as Witness in Victorian Science* (Johns Hopkins Univ. Press, 2005).
Jones, C. & Galison, P. (eds) *Picturing Science Reproducing Art* (Routledge, 1998).
Interactive website of Cambridge's Whipple Museum for the History of Science: www.hps.cam.ac.uk/whipple/explorecollections

# BOOKS & ARTS

# Keeping up scientific standards

A journalistic account of the case of data manipulation by physicist Jan Hendrik Schön is rich in detail but draws the wrong conclusions about the self-correcting processes of science, argues **Martin Blume**.

**Plastic Fantastic: How the Biggest Fraud in Physics Shook the Scientific World**
by Eugenie Samuel Reich
Palgrave Macmillan: 2009.
272 pp. £15.99, $26.95



Jan Hendrik Schön's case raised the need to educate researchers about ethics.

Jan Hendrik Schön first came to New Jersey's Bell Laboratories from his doctoral work at the University of Konstanz, Germany, as an intern in 1997. A year later he became a postdoc and staff member. Over the next few years he published apparently groundbreaking research in materials science. His results appeared in major research journals, and his co-authors and supervisors included highly respected researchers. Scientists worldwide tried to replicate his findings — but were unable to do so. Mounting complaints about inconsistencies in his work led to allegations of scientific misconduct being made against Schön. He was investigated by the management at Bell Labs, first internally and then by a select external committee in 2002. This committee found compelling evidence of manipulation and misrepresentation of data. Schön was fired from his position and has not worked visibly in research since.

Journalist Eugenie Samuel Reich pursues this affair in depth in *Plastic Fantastic*. Holder of a bachelor's degree in physics and philosophy from the University of Oxford, UK, Reich is fascinated by scientific conduct. Her investigation is the result of several years of research, during which she interviewed 125 individuals who interacted with Schön, attended meetings and visited the laboratories where attempts were made to replicate his 'discoveries'.

The introduction and first chapter can be understood by any reader. Reich explains the strange title of the book — "plastic fantastic" was a phrase used by the media to describe Schön's claims of finding superconductivity in a plastic material, and it played into her quest to contact him. She discusses how her interest in the story arose from her reading of "the most riveting piece of technical writing I have ever come across" — namely the final investigation report into Schön's conduct by the external committee chaired by Malcolm Beasley, a physics professor at Stanford University in California.

Later chapters of the book place Schön's work in context; basic research at Bell Labs was in decline at that time because of restructuring, and its researchers were under pressure. Reich also focuses on the peer-review process at the journals *Science* and *Nature*, in which Schön published many prominent articles. She describes the science behind the recondite fields in which Schön worked, including 'MOSFET' electronic chips and superconductivity in plastics. Here the writing becomes too technical and the book's layout, with some 200 footnotes, distracting. Surprisingly there are no acknowledgements. Although some interviewees are mentioned, we do not learn which experts' views contributed the most. Nevertheless, the book will interest physicists and materials scientists working in related fields, and those who strive to prevent scientific misconduct.

Reich does an excellent job of dealing with the facts of the Schön case, but less well with their interpretation. She tends to describe issues in black-and-white terms and uses strident language unnecessarily. An acerbic tone creeps into judgements of the individuals involved, such as one editor being "opinionated", and this distracts from her central points.

More important is Reich's conclusion that the self-correction process of science failed in this case and should not be trusted: "It seems like little more than blind faith to insist that all

**"The possibility of scientific misconduct cannot be eradicated."**

activity carried out in the name of science will always be self-correcting." Yet Reich doesn't say where this assumption comes from. No scientist who has thought about how science works would "insist" on this; nor can it "always" be true. It is likely that fabrication of data in a significant discovery will be uncovered because close attention will be paid to it. But fabricated data in a more obscure article may not be corrected.

Reich understands how data fabrication may arise. She admits that she made up experimental data while a student working as a research assistant. Seeing a pattern emerging in her measurements, she entered values that fitted a trend rather than recording objectively what she saw. Her supervisor did not notice. Reich remarks that she had not been educated on the sanctity of data beforehand, only realizing its importance after recalling the incident when writing this book. Harm might have been done if she had missed a change in the pattern that signalled a new discovery. This example of a student seeking the expected answer rather than new clues shows how important it is to educate researchers. Reich's late recognition of this issue should also have made her pause before drawing strong conclusions on the processes of science.

Although the detection of data fabrication is not inevitable, we can increase the probability of detection by remaining vigilant at each stage in the scientific process. Steps include attempting to replicate important results and encouraging the alertness of managers, journal editors and referees, and of knowledgeable scientists who read and evaluate publications. In the Schön case, two scientists in particular played a key part — Paul McEuen of Cornell University, New York, and Lydia Sohn, then of Princeton University, New Jersey. Combing through Schön's papers, they found inconsistent and repeated figures that pointed to serious problems, which were brought to the attention of Bell management. Reich sees the need of individuals to intervene as illustrating science's failure to cope with misconduct; but their vigilance is part of the correction process, which culminated in the Beasley report that Reich so admires.

The problems with Schön's work should have been detected earlier. Although his co-authors

and supervisors were all cleared of misconduct in the final report, they could have been more curious about his results. Similarly, when Schön applied for patents, he was not required to show or have others sign his laboratory notebook — his use of which is questionable.

It must also be noted that an accusation of misconduct is not a finding. And an incorrect or false accusation can ruin reputations or destroy careers, even if the accusation is later recognized as wrong. This puts a burden on institutions that requires thought and investigation before public disclosure, which slows the process.

Reich opines that the memory of this particular affair will fade. But the possibility of scientific misconduct cannot be eradicated. Many people are involved in educating scientists at all stages of their careers to keep the possibility of scientific fraud, plagiarism, conflict of interest and other unethical behaviour before them. For example, at a 2003 workshop that I convened in London, more than 70 international participants, including publishers, editors and working scientists, discussed aspects of unethical behaviour, as well as the Schön affair. The result was a set of guidelines that was adopted, after minor modification, by the council of the International Union of Pure and Applied Physics (see http://tinyurl.com/iupap-ethics).

Scientific associations, such as the American Physical Society (APS) and the Council of Science Editors, have developed and updated codes of behaviour and educational programmes before and since this landmark case. The APS policy (http://tinyurl.com/aps-policy) includes a report on ethics education, which is vital to define and promote ethical behaviour by all scientists. Reich herself references the Beasley report using a link on the website of the APS. I was responsible for posting it there, with permission from Bell Labs' owners Lucent Technologies, following the retraction of six articles co-authored by Schön in the APS journals for which I was then editor-in-chief. It is fortunate that this stable link keeps the report accessible years later (see http://publish.aps.org/reports).

The prevention and uncovering of scientific misconduct requires vigilance, education and humility on the part of scientists, managers, editors, journals and journalists to maximize the probability of correction. In the Schön affair, several of these elements combined to raise that probability enough so that correction occurred. It is unfortunate that Reich did not draw fairer conclusions, despite her hard work in reporting the facts. ■

**Martin Blume** is editor-in-chief emeritus at the American Physical Society and senior physicist emeritus at Brookhaven National Laboratory, Upton, New York, USA.
e-mail: blume@post.harvard.edu



Buddha is said to have wooed his future wife by reeling off a huge number series.
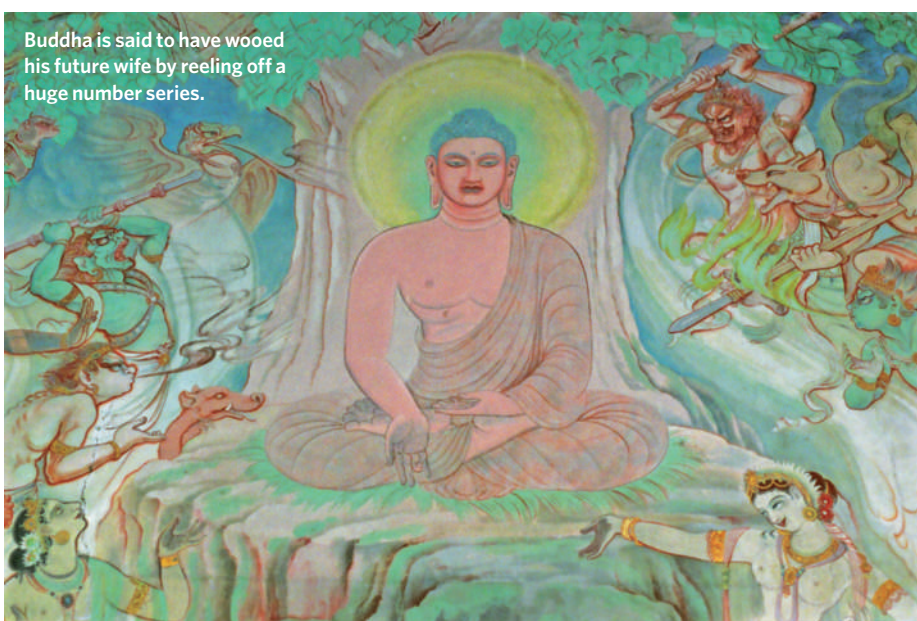
F. SOLTAN/SYGMA/CORBIS

# An Indian history of numbers

In a world divided by culture, politics, religion and race, it is a relief to know one thing that stands above them — mathematics. The diversity among today's mathematicians shows that it scarcely matters who invents concepts or proves theorems; cold logic is immune to prejudice, whim and historical accident. And yet, throughout history, different families of humans have distilled the essence of the cosmos to capture the magic of numbers in many ways.

*Mathematics in India* shows just how different one of these ways was, and how culture and mathematical development are intimately connected. This carefully researched chronicle of the principal contributions made by a great civilization covers the earliest days of Indian history through to the beginning of the modern period. Regrettably, it stops short of the legendary mathematician Srinivasa Ramanujan (born 1887), whose name is still seen in today's research papers.

Kim Plofker's book fulfils an important need in a world where mathematical historiography has been shaped by the dominance of the Greco-Christian view and the Enlightenment period. Too little has been written on the mathematical contributions of other cultures. One reason for the neglect of Indian mathematics was Eurocentrism — British colonial historians paid it little attention, assuming that Indians had been too preoccupied with spiritual matters to make significant contributions to the exact

> **Mathematics In India**
> by Kim Plofker
> Princeton University Press: 2009.
> 394 pp. £28.95

sciences. Another reason is that many ancient Indian mathematical texts have long been extinct; often, the only indication that they existed comes from scholars who refer to the work of their predecessors. As Plofker wryly notes, two historians of Indian maths recently published articles in the same edited volume, wherein the estimates of their subject's origins differed by about 2,000 years.

Still, surviving Sanskrit texts reveal a rich tradition of Indian mathematical discoveries lasting more than 2,500 years. In the Early Vedic period (1200–600 BC), a decimal system of numbers was already established in India, together with rules for arithmetical operations (*ganita*) and geometry (*rekha-ganita*). These were encoded in a complex system of chants, prayers, hymns, curses, charms and other religious rituals. Cryptic phrases called *sutras* contained arithmetical rules for activities such as laying out a temple or arranging a sequence of sacrificial fires.

Large numbers held immense fascination. Acclamations of praise to the air, sky, times of day or heavenly bodies were expressed in powers of ten that went to a trillion or more. Reputedly, the young Prince Buddha successfully competed for the hand of Princess Gopa by reciting a number table that included names for the powers of ten beyond the twentieth decimal place.

As in other early agricultural civilizations, Indian mathematics probably emerged in

response to the need to measure land areas and keep track of financial transactions, incomes and taxation. A rigid caste and class hierarchy reserved the mystery of numbers for elite Brahmins. To maintain personal power, mathematical knowledge was jealously guarded. Its communication was deliberately made difficult, such as in the perplexing rhythmic chant of mathematician Aryabhatta in the fifth century AD: "*makhi-bhakhi-phakhi-dhaki-nakhi-nakhi-nakhi-hasjha-skaki-kisga-sghaki-kighva-ghaki…*" This recital of values of sine differences in arc minutes would be memorized by aspiring mathematicians in much the same way as verses of the sacred text Bhagavadgita.

The book details the impressive achievements of Indian mathematicians, from Aryabhatta through Brahmagupta, Mahavira, Bhaskara and Madhava, until the Sanskrit tradition became irrelevant with the invasion of modern mathematics from Europe in the nineteenth century. Major discoveries include finding the solution to indeterminate equations and the development of infinite series for trigonometric quantities. Discovered in the fourteenth century by the Kerala school founded by Madhava, these series built on the work of Bhaskara II and grew from the ingenious computation of a circle's circumference. By breaking up the circle into polygons, Madhava was able to calculate the value of pi correct to 11 decimal places. Some developments preceded those in Europe. For example, Reuben Burrow — a British mathematician posted to Bengal as an instructor in the engineers corps — was intrigued by rules he discovered in an unnamed Sanskrit text, and wrote a paper in 1790 entitled 'A Proof that the Hindoos had the Binomial Theorem'.

But how peculiarly Indian was early Indian mathematics? Did it evolve in isolation or did it absorb ideas and knowledge from elsewhere? Cultural pride in their recently reinvigorated country causes some Indians to claim that all worthwhile mathematics originated in ancient India. But this book will not please them. Plofker is not ready to certify that the concept of zero was an Indian invention; it could well have been conveyed by Chinese Buddhist pilgrims. Nor is she willing to believe that differential and integral calculus were anticipated in India ahead of the work of Gottfried Leibniz and Isaac Newton.

The chapter entitled 'Exchanges with the Islamic World' is of particular significance. The Muslim conquest of India brought with it the Islamic mathematical tradition, which was founded on Greek mathematics. Muslims made important advances in maths between the ninth and thirteenth centuries. Greco-Islamic and Indian mathematics were structured quite differently, with the former emphasizing proof and the latter, result. Probably because of Islamic influence, Indian ideas on the nature of mathematical proof moved in the direction of greater rigour.

The book carefully separates fact from hyperbole, copiously quoting formulae. This makes for heavy reading in places, and one wishes that it had been interspersed with vignettes and light anecdotes. It is more of a research monograph than a popular book. But that is the price that scholarship exacts.

*Mathematics in India* explains how the early development of Indian maths was influenced by religion, by the need to build temples of specific proportions and to meet astrological imperatives. Similarly, it could be argued that Islamic mathematics was religiously motivated — for example, by the need to know the precise times of daily prayers, and to determine the direction of the holy *Kaaba* (the *Qibla*). But a quadratic equation solved by whoever, by whatever means and for whatever purpose must give exactly the same solutions. Ultimately, mathematics is mathematics. ■

**Pervez Hoodbhoy** is professor of physics at Quaid-e-Azam University, Islamabad, Pakistan.
e-mail: pervez.hoodbhoy@gmail.com

# Averting environmental crisis

One of the most important questions of our age is when will humankind take action to offset the impact of climate change? Will we do something before the inexorable rise in sea level or wait until the effect is cataclysmic? Can the history of previous environmental crises help us to predict the course of this one?

Two books address these challenges. In the first, *Becoming Good Ancestors*, David Ehrenfeld, a zoologist and the founding editor of the journal *Conservation Biology*, examines the destructive tendencies of humankind. He asks if we can "move ourselves and our society toward a more stable, less frantic, more responsible, and far more satisfying life". The book is an expanded and revised collection of some three dozen essays that Ehrenfeld published first in *Orion* magazine and then in *Swimming Lessons: Keeping Afloat in the Age of Technology* (Oxford University Press, 2002). The result is a very good read.

Ehrenfeld believes that we must jettison our arrogant assumption of being able to fix anything through technology, that we forget at our peril what worked well in the past and that economics is a cloudy lens through which to view human behaviour. He says that we must and can reconnect with a nature that is resilient and that, despite globalization, local communities will never completely disappear. The essays retain the qualities that made them appealing when they were first published — brevity, passion and accessibility. However, as satisfying as they are as self-contained meditations, they do not hang together well, either as a sequence or as a complete analysis.

**Becoming Good Ancestors: How We Balance Nature, Community, and Technology**
by David Ehrenfeld
Oxford University Press: 2008.
320 pp. $19.95, £10.99

**Nature and Power: A Global History of the Environment**
by Joachim Radkau
Cambridge University Press: 2008.
448 pp. $24.99, £14.99

*Nature and Power* is a very different work, written by an environmental historian who refuses to follow convention unless he finds compelling reason to do so. Joachim Radkau's vision in this broad-reaching history of the state of the environment — in particular, its soils, forests and waters — is from three perspectives, namely German, continental European and global. His unrelenting focus on detail may frustrate the reader who seeks straightforward narrative. Many writers can be said to miss the forest for the trees, but Radkau is extreme, at times abandoning the tree for the twig, bud, leaf or abscission scar.

Radkau's guiding lights are demographer Thomas Malthus, sociologist Max Weber and the nineteenth-century agricultural chemist Justus von Liebig. Radkau draws on von Liebig in his discussions on the chemistry, well-being and degradation of soils; on Weber in his acknowledgement of the crucial role of culture and power in the course of environmental history; and on Malthus for his case that population pressure on resources is the enemy of sustainability. Radkau doesn't just apply these and many other ideas, but tests, refines and refutes them in chapters that range through time and across the globe. The book often startles. Insights come thick and fast with Radkau's ironic and unexpected turns of phrase.

True to form, he refuses to take the easy route. He comes close to declaring that environmental history is mostly about decline, and that human population control is the key to reducing resource pressure, yet he pulls back from whatever precipice he is nearing

and complicates his narrative with stories that either offset or flatly contradict his thesis. Rather than close a case with some glib conclusion, he reminds the reader that the course of environmental history is intertwined with human power and inertia, that it is a mix of decline, ascension and stability and that crisis is often contrived. Radkau keeps the reader off balance: "All simple pictures of environmental history are open to challenge."

Rarely, it seems, have we acted to prevent environmental crises. An exception was the 1987 Montreal Protocol to cease production of chlorofluorocarbons and other compounds that deplete the ozone layer. Will concern over climate change — a greater, more complex and more diffuse issue than destruction of ozone — also produce a pact for change? Given the cost of ramping up such efforts, this seems unlikely before climate change becomes a

worldwide cataclysm. Whereas Ehrenfeld may give you cause to hope, Radkau is likely to leave you gloomy. He might even be said to turn philosopher George Santayana's observation on its head: even those who remember the past seem condemned to repeat it. ∎

**Shepard Krech III** is a professor in the Department of Anthropology at Brown University, Providence, Rhode Island 02912, USA.
e-mail: krech@brown.edu

# Q&A: Acting the part

Actor and playwright **Anna Deavere Smith** has pioneered documentary theatre through her one-woman plays constructed from interviews. As she prepares to portray biologists Edward O. Wilson and James Watson at the World Science Festival in New York next week, Smith talks about life, death and the influence of science on her work.

**Watching Wilson and Watson**
Skirball Center for the Performing Arts, New York City
11 June 2009. Part of the World Science Festival, 10–14 June.

**When did you start writing plays based on interviews with real people?**
Nearly 30 years ago, I created a project called *On The Road: A Search for American Character*. The idea was to travel the country with a tape recorder, talk to people and invite them to see themselves performed. Originally it was going to be with a company of actors, but I thought I'd do all the parts until I figured out how to raise money to pay them. As a kid I was a mimic. I find my expression through others.

**How do you get people to open up to you?**
In the early days I wanted people to talk to me in individual, peculiar ways. A linguist gave me three questions to ask to ensure that would happen: have you ever come close to death? Have you ever been accused of something you didn't do? And do you remember the circumstances of your birth?

**When did you get interested in science?**
As a child, I wanted to be a psychiatrist, or an inventor like Thomas Edison. I admire scientists' sense of experimentation, their tolerance for not knowing. While teaching drama at Stanford University, California, in the 1990s, I became infatuated with a geneticist there, Marcus Feldman, who studies evolution using twins. He told me he'd spent ten years of his career trying to shoot down the views of William Shockley



Playing solo: Anna Deavere Smith.

and Arthur Jensen, who argued that intelligence varies by race. Feldman became a muse for me.

**Why did you decide to impersonate Edward O. Wilson and James Watson?**
Watson is one of the forces behind the celebration of Wilson's 80th birthday at the World Science Festival this year. He invited me to do a 20-minute performance of each of them. When they were young scientists at Harvard University, there was a rift between them. Watson wouldn't speak to Wilson, and Wilson later wrote that he had thought Watson was "the most unpleasant human being [he] had ever met". They've reconciled over the years.

**What is Wilson like?**
He's fashioned himself as a southern gentleman: very friendly, patient, charming and with a ready smile. He was a boy scout,

and in some ways he's still a grown-up boy, in that he has that restlessness and excitement about learning something new. He developed a work ethic as a child when he had to get up at 3 a.m. for a paper route, and he now gets up very early with great purpose. Just look at all those big fat books he has written.

**Why did you choose the topic of health care for your next play, which opens in New York City later this year?**
*Let Me Down Easy* is about the beauty of life and the fact that it has an expiration date. In the late 1990s, Yale University School of Medicine asked me to interview doctors and patients and portray them at medical rounds. Since then, I haven't had the desire to make a play about anything else. The project has expanded in my mind from medicine to a long excursion into the human body, its resilience and vulnerability. The play ranges from portraying people who have physical prowess, such as cyclist Lance Armstrong and long-distance swimmer Lynne Cox, to people who are dying for no reason other than chance. And I too have had to come to grips with the fact that I'm going to die.

**Has your work raised questions for you?**
It has left me with unanswered questions about the relationship between speech and inner life. I don't understand exactly what happens when a word enters my imagination, or when I reiterate the word as it was said. There is probably a psychologist, neurologist or linguist who would offer a lot to my study. I should talk to some experts. I'm the machine but I don't know entirely how it works. ∎
Interview by **Jascha Hoffman**, a writer based in New York.

A.H. WALKER/GETTY

# NEWS & VIEWS

**PHOTONICS**

# How nanocrystals lost their blink

Taekjip Ha

**Semiconducting nanocrystals emit light in many different colours, but they blink on and off at random. The latest nanocrystals emit photons steadily, thanks to the blending of their cores into their outer shells.**

High on the wish list of many scientists is a light source that is very small and bright, and that can be turned on at will. Possible applications abound, from ultra-high-resolution imaging of single proteins in cancer cells to high-performance lasers and light-emitting diodes. For two decades, semiconductor nanocrystals have been heralded as that dream light source. And why not? They are only a few nanometres in diameter, are about ten times brighter than the best organic dyes and can be made in many different sharply defined colours, which is useful for multiplexing. Furthermore, they degrade much less easily in light than other nanometre-scale light sources.

Alas, these nanocrystals have a well-known problem: they do not emit a constant stream of light. Instead, they turn on and off randomly[1] and on no particular timescale[2]. Using these 'blinking' particles as light sources is akin to using stroboscopic illumination in everyday situations — it might be fun on a dance floor, but it's not much good for lighting a movie set if you want film-goers to see what's actually happening. It is therefore highly desirable to eliminate blinking from nanocrystals and, unsurprisingly, many researchers have attempted to do just that for 15 years. Reporting on page 686 of this issue, Wang *et al.*[3] describe an unqualified success in producing nanocrystals that don't blink at all.

So what's the trick? Conventional semiconductor nanocrystals, also known as quantum dots, consist of a core of one material surrounded by a shell of another[4,5] (Fig. 1a, b). Typically, the core is made up of two elements, such as cadmium and selenium atoms, whereas the shell is often a mixture of zinc and sulphur atoms. If a photon excites the core, an electron–hole pair is created (where a hole is a notional particle created by the absence of an electron). The shell prevents the pair from spreading into the surroundings, thereby ensuring that photoexcitation results in the emission of a photon, rather than the generation of heat.

But in Wang and colleagues' recipe[3] for quantum dots, the core is made up of three types of atom — cadmium, selenium and zinc. Furthermore, the boundary between the core and the shell is not abrupt, as in traditional
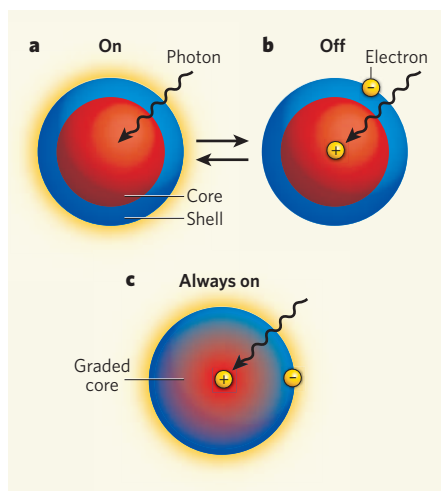


**Figure 1 | Light emission from nanocrystals. a**, Conventional nanocrystals typically consist of a core of one semiconducting material surrounded by a shell of another. When a photon excites the core, light is emitted from the nanocrystal as it relaxes back to the ground state. **b**, Occasionally, an electron escapes the core and is trapped on a surface defect on the shell, leaving the core positively charged. The charged core can still be excited by light, but the excitation dissipates as heat. The nanocrystal thus stops emitting photons until the trapped electron returns to neutralize the core. **c**, Wang *et al.*[3] have prepared nanocrystals in which the core and the shell blend into each other. This seems to change the relaxation mechanism of the nanocrystals so that they emit photons efficiently even when charged.

quantum dots, but graded, so that the materials of the core and shell blend into each other (Fig. 1c). Amazingly, when the authors embedded these 'alloyed' particles in a polymer matrix and examined them under a fluorescence microscope, they saw that individual particles emitted photons at a constant level for hours, without blinking once.

Blinking is so distinctive of single nanocrystals that its detection is often used to prove that such crystals have been prepared as isolated particles (rather than as aggregates, which don't blink). As a result, Wang *et al.* had to use a sophisticated method to demonstrate that

their non-blinking objects are indeed single nanocrystals. Their approach depended on the fact that single nanocrystals cannot emit two photons simultaneously; any two successive photons must therefore be separated by at least the lifetime of an emission, a phenomenon known as photon antibunching. The authors' nanocrystals displayed beautiful antibunching, with each emission lasting about 5 nanoseconds — a much shorter duration than that of conventional nanocrystals, hinting at an unusual emission mechanism for the new quantum dots.

Why did it take so long to create non-blinking nanocrystals? It certainly wasn't for lack of trying. Hundreds of papers have been written about the mechanism of blinking since the first report of the phenomenon[1], many of which were published as a result of failed attempts to suppress it. The motto has been: if we can't get rid of it, let's study it! There is now a consensus that a nanocrystal stops emitting photons when it becomes charged, probably when electrons ejected from the core become trapped in defects on the nanocrystal's surface. A charged nanocrystal can still be excited by photons to produce an electron–hole pair, but this excitation is lost as heat (rather than by photon emission) in a process termed Auger recombination. Charged nanocrystals therefore remain dark until surface-trapped electrons return to the core, whereupon photon emission resumes (Fig. 1a).

If this model is correct, then there are three strategies for curing blinking. The first is to eliminate any surface traps (or to modify them to stop them from ensnaring electrons). Traps have been eliminated using innovations in nanoparticle synthesis[6,7] or rendered inactive by chemical treatment[8], but in all cases, although impressively reduced, blinking was not prevented. The second approach is to find a way of returning trapped electrons to the core as quickly as possible. There is currently no reported use of this method.

The final strategy is to find a way to maintain photon emission even when nanocrystals are charged (Fig. 1c). Wang and colleagues' theoretical analysis[3] of their system indicates that this is what happens in their quantum dots: the graded core–shell boundary makes

Auger recombination inefficient, while also enabling interplay between the positive charge on the nanoparticle and the photo-generated electron–hole pairs. This interplay gives rise to luminescent signals unhindered by any competing heat-generation processes. The proposed mechanism also helps to explain some unusual features of the nanoparticles' emission spectra, such as the fact that they are broad and contain several peaks.

This work is a milestone in the rapidly growing field of colloidal nanomaterials, and raises several questions. Are the nanocrystals constantly charged, or do they switch back and forth between uncharged and charged states without causing any change in signal intensity[9]? The nanocrystals' broad, multi-peaked spectra will limit potential biological applications that require multicolour imaging[10], so is it possible to develop non-blinking nanocrystals that retain the narrow emission spectra of other quantum dots? And finally, can Wang and colleagues' nanocrystals be adapted for biological imaging? This would require the nanocrystals to be further engineered to make them water soluble, and so that they can be attached to biological molecules without sacrificing their non-blinking properties. On the basis of experience with other quantum dots, the desirable optical properties of the alloyed nanocrystals are likely to be maintained even after further processing.

The discovery of fluorescent blinking has been one of the most striking discoveries of single-molecule spectroscopy. Nanocrystals do it. Green fluorescent proteins do it[11]. Even certain polymers[12] and organic dyes do it[13]. Blinking seemed to be intrinsic and unavoidable for all single emitters exposed to the elements — until now. Wang and colleagues' discovery[3] is therefore cause for a small celebration, before researchers delve further into the underlying cause and seek possible applications. ∎

Taekjip Ha is in the Department of Physics and the Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign, Howard Hughes Medical Institute, Urbana, Illinois 61801, USA.
e-mail: tjha@illinois.edu

1. Nirmal, M. et al. Nature 383, 802–804 (1996).
2. Kuno, M., Fromm, D. P., Hamann, H. F., Gallagher, A. & Nesbitt, D. J. J. Chem. Phys. 112, 3117–3120 (2000).
3. Wang, X. et al. Nature 459, 686–689 (2009).
4. Peng, X., Schlamp, M. C., Kadavanich, A. V. & Alivisatos, A. P. J. Am. Chem. Soc. 119, 7019–7029 (1997).
5. Dabbousi, B. O. et al. J. Phys. Chem. B 101, 9463–9475 (1997).
6. Chen, Y. et al. J. Am. Chem. Soc. 130, 5026–5027 (2008).
7. Mahler, B. et al. Nature Mater. 7, 659–664 (2008).
8. Hohng, S. & Ha, T. J. Am. Chem. Soc. 126, 1324–1325 (2004).
9. Shimizu, K. T., Woo, W. K., Fisher, B. R., Eisler, H. J. & Bawendi, M. G. Phys. Rev. Lett. 89, 117401 (2002).
10. Michalet, X. et al. Science 307, 538–544 (2005).
11. Dickson, R. M., Cubitt, A. B., Tsien, R. Y. & Moerner, W. E. Nature 388, 355–358 (1997).
12. Vanden Bout, D. A. et al. Science 277, 1074–1077 (1997).
13. Ha, T. et al. Chem. Phys. Lett. 271, 1–5 (1997).

## COSMOLOGY

# Climbing up the cosmic ladder

Ofer Lahav

**Study of an under-appreciated subclass of stars that brighten and dim periodically brings to light their potential use as cosmic yardsticks — out to distances of a few hundred million light years.**

What is the expansion rate of the Universe? And how old is the Universe? The answers to these and other fundamental cosmological questions depend on a number of factors, among them the measurement of the Hubble constant ($H_0$). This is defined as the ratio of a galaxy's recession velocity — that is, the speed at which it is moving away from Earth — to its distance. Writing in *The Astrophysical Journal*, Bird et al.[1] set out to study a class of stars that may well assist in measuring $H_0$ and so help to answer these questions.

Although it is relatively easy to measure a galaxy's recession velocity from the amount by which emission and absorption lines in the galaxy's spectrum are shifted towards redder wavelengths, determining its distance from Earth is much more challenging. But nature has been kind enough to provide astronomers with stars that act as 'standard candles': if they know how bright the candles are, and how bright they look from Earth, they can

calculate how far away they are. The apparent flux of a candle is simply proportional to its intrinsic luminosity divided by the square of its distance. Therefore, provided that the intrinsic luminosity of the standard candle is well calibrated, researchers can measure the distance to its host galaxy.

Heroic efforts have been made to identify standard candles. One class of stars, the Cepheids, has become particularly popular. Named after their prototype, Delta Cephei, Cepheids are variable stars marked by a tight correlation between their period of variability and their intrinsic luminosity: the greater the period of the star, the greater its maximum brightness. This period–luminosity (P–L) correlation, discovered about 100 years ago by Henrietta Leavitt, makes Cepheids very useful standard candles.

But using Cepheids as cosmic-distance indicators requires their intrinsic luminosities to be calibrated through distance measurements

to their nearby counterparts on the basis of methods other than the P–L correlation. Hence the term 'cosmic ladder': each distance measurement relies on calibration by nearer objects. In their study, Bird et al.[1] propose a way of measuring Cepheid distances out to 100 megaparsecs (1 Mpc = 3.26 million light years) and beyond, greatly improving the reliability of the cosmic ladder.

The value of the Hubble constant has been a subject of debate for decades, most notably between Allan Sandage and Gérard de Vaucouleurs, who argued for values of 50 and 100 km s$^{-1}$ Mpc$^{-1}$, respectively. By analysing Cepheids observed with the Hubble Space Telescope, Wendy Freedman and colleagues[2] have derived a value of $H_0 = 72 \pm 8$ km s$^{-1}$ Mpc$^{-1}$, and a very recent study of 240 Cepheids observed with an improved camera onboard the Hubble Space Telescope has yielded an even more accurate figure, $74.2 \pm 3.6$ km s$^{-1}$ Mpc$^{-1}$ (ref. 3). What is remarkable is that these values agree so well with those deduced from studies that combine observations of the cosmic microwave background radiation (relic radiation from the Big Bang) carried out with the Wilkinson Microwave Anisotropy Probe with other cosmological data: $70.5 \pm 1.3$ km s$^{-1}$ Mpc$^{-1}$ (the age of the Universe inferred from such studies is $13.72 \pm 0.12$ billion years)[4].

However, the case is by no means closed: there are persistent claims for lower values, such as $62.3 \pm 1.3$ km s$^{-1}$ Mpc$^{-1}$ (ref. 5). Only time will tell whether the currently endorsed value of about 70 km s$^{-1}$ Mpc$^{-1}$ will survive further scrutiny. Future research, using better and bigger Cepheid samples, is necessary to reassess its value and increase Cepheids' reliability as cosmic-distance indicators.

The downside of conventional (short-period) Cepheids is that they only 'climb' the cosmic ladder out to distances of about 20 Mpc. More distant Cepheids would be helpful both in extending the ladder and in calibrating studies that rely on another class of more distant standard candles — type Ia supernovae — to determine the expansion rate of the Universe.

Enter Bird and colleagues[1]. They draw attention, by re-analysing data from the literature, to a somewhat neglected class of objects called ultra-long-period (ULP) Cepheids. They re-examine 18 Cepheids that have periods ranging from 80 to 210 days. As the P–L relationship dictates, if their periods are longer, they are brighter and so can be observed over greater distances — 100 Mpc and beyond. As such, these Cepheids offer the potential to improve the accuracy of connecting the steps — or even skipping unnecessary steps — of the cosmic ladder.

The authors[1] also note that these ULP Cepheids have a relatively 'shallow' P–L relationship — the luminosity changes modestly with the period. They conclude that ULP Cepheids are more standard-candle-like than classical Cepheids, in particular when observed at wavelengths for which extinction due to dust

is smaller. They also claim that the uncertainty in the derived cosmic distances is as good as that obtained with classical Cepheids or type Ia supernovae.

But as the authors themselves acknowledge, ULP Cepheids' use as distance indicators also comes with limitations. First, they are relatively rare. Second, they require more observations to cover their long variability periods. Future observations, ideally undertaken with telescopes in space, would both help to overcome this limitation and offer insight into the evolution of these stars.

What Bird *et al.* did not attempt to do was test the implications of the observed ULP Cepheids for the value of $H_0$. In the wider picture, a more accurate determination of Cepheid and supernovae distances will have profound implications for understanding the basic properties of our Universe. The Hubble

constant is just the value of the Hubble function $H(z)$ at the present epoch (or redshift $z = 0$). Cosmologists would like to measure $H(z)$ accurately out to high redshifts, because this function depends on the contents of the Universe.

The current measurements imply that only about 4% of the Universe is made up of normal (baryonic) matter, whereas the rest of it is made up of cold dark matter (21%) and the even more mysterious dark energy (75%), which seems to be causing the expansion of the Universe to speed up. But other scenarios, such as an epoch-dependent dark energy or modifications to Einstein's theory of general relativity, are not excluded by the observations.

There are also indirect methods of estimating $H(z)$, such as those based on the large-scale distribution of galaxies, clusters of galaxies and gravitational lensing (the deflection of light from distant sources by foreground

objects). And indeed, there are plans for several dark-energy surveys of billions of galaxies that will lead to a more refined estimation of $H(z)$ and other cosmological functions. Better calibrated samples of type Ia supernovae will also be crucial for pinning down $H(z)$, and the proposed class of ULP Cepheids may well play an important part in this effort. ∎

Ofer Lahav is in the Department of Physics and Astronomy, University College London, London WC1E 6BT, UK.
e-mail: lahav@star.ucl.ac.uk

1. Bird, J. C., Stanek, K. Z. & Prieto, J. L. *Astrophys. J.* **695,** 874–882 (2009).
2. Freedman, W. L. *et al. Astrophys. J.* **553,** 47–72 (2001).
3. Riess, A. G. *et al.* preprint at http://xxx.lanl.gov/abs/0905.0695 (2009).
4. Komatsu, E. *et al. Astrophys. J. Suppl. Ser.* **180,** 330–376 (2009).
5. Tammann, G. A., Sandage, A. & Reindl, B. *Astron. Astrophys. Rev.* **15,** 289–331 (2008).

---

## MEMBRANE-PROTEIN STRUCTURE

# Piercing insights

## Hagan Bayley

**Pore-forming proteins are deadly biological weapons that punch holes in target-cell membranes. The structure of the pore formed by a bacterial toxin suggests that diverse pore formers have similar assembly pathways.**

Nature's pore-forming proteins are primitive but lethal weapons[1]. Bacteria use them to wage war against rival bacteria and to attack human cells. Not to be outdone, our immune defences use pore formers too, such as the complement component C9, which attacks bacteria and protozoa, and perforin, a protein that kills virus-

infected cells. But despite their ubiquity, very little is known about the detailed structures of the pores formed by these proteins, as their elucidation presents substantial technical challenges. Fortunately, on page 726 of this issue, Mueller and colleagues[2] now describe a striking X-ray structure of the pore formed by a bacterial

toxin, which differs markedly from the sole previous example[3]. Nevertheless, they propose an assembly mechanism that has features in common with many other pore formers.

Many pore-forming proteins damage their targets simply by forming holes in the targets' outer membranes, whereas others, such as anthrax toxin, transport noxious substances into the victim's cells. Unlike most membrane proteins, which are inserted into membranes within the cell as they are made, pore-forming proteins are usually made in water-soluble, secreted forms, which then assemble in the target-cell membrane.

The pores derived from soluble pore-forming proteins fall into two main classes. The best described are pores with transmembrane β-barrels, in which polypeptide strands weave back and forth across the lipid bilayer, creating a sheet that forms a cylindrical surface. One example is the α-haemolysin of *Staphylococcus aureus*, which spans the lipid bilayer of target cells as a 14-stranded barrel, with two polypeptide strands contributed by each of seven subunits[3]. By contrast, each subunit of the large pores formed by the extensive family of cholesterol-dependent cytolysins (CDCs) contributes four strands to each membrane-spanning β-barrel, which comprises up to 50 subunits[4].

In the second class, there are many pores that are thought to be formed from spiral polypeptide rods, known as α-helices, rather than from extended β-strands[1]. These proteins are often used in bacterial assaults on other bacteria. Despite considerable effort, attempts to obtain three-dimensional structures of α-helical pores, either by X-ray diffraction or at lower resolution by electron microscopy, have failed. Most of these proteins probably resemble antimicrobial peptides, which often form poorly defined or weakly associated oligomers, militating against structure determination. Therefore, the structure of the α-helical toxin pore formed by cytolysin A (ClyA) described by Mueller
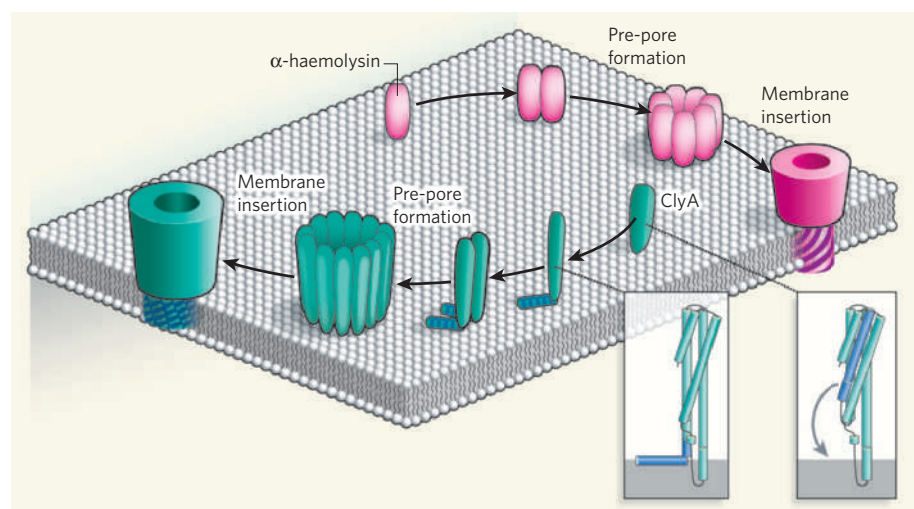


**Figure 1 | Assembly of staphylococcal α-haemolysin and ClyA pores show common principles.**
The staphylococcal α-haemolysin pore (pink) is made up of seven subunits, which span the lipid bilayer as a β-barrel, whereas ClyA (green) forms an iris-like barrel of α-helices from its twelve subunits[2]. The conformational changes undertaken by α-haemolysin before insertion are minimal; by contrast, ClyA undergoes radical reorganization before membrane penetration, with extension of one of the three helices in each subunit in an arc towards the membrane surface (as shown in the insets). Even so, crucial aspects of the assembly pathway, including the participation of a non-penetrating pre-pore intermediate, are similar.

**NEWS & VIEWS**

and colleagues[2] is exceptionally enlightening. Furthermore, the authors use a previously determined structure of the ClyA monomer[5] to speculate about the assembly pathway.

ClyA is produced by pathogenic strains of *Escherichia coli* and *Salmonella enterica*. The ClyA pore is formed from 12 subunits (Fig. 1), each of which contains three long helices. The structure predicts that the tips of two helices partially span the bilayer, with the third spanning it completely. The authors[2] maintain that, to attain its imposing oligomeric structure, monomeric ClyA contains a spring-like device that is activated on contact with the target membrane. First, the hydrophobic tip of the molecule, which in earlier studies was thought to be the only domain involved in membrane penetration, contacts the lipid bilayer. This results in activation of the spring to extend the first two helices, allowing them to enter the bilayer. A second step, unanticipated by earlier work, is the nunchaku-to-staff extension of the third helix, in which its tip moves through 140 ångstroms, swinging in an arc towards the membrane surface (Fig. 1, inset). This extraordinary lengthening of the protein has at least one precedent in nature — the enormous conformational change that occurs in influenza virus haemagglutinin when viral- and host-cell membranes fuse during infection[6].

After the second step, the top end of the elongated third helix remains incompletely extended and lies parallel to the membrane surface (Fig. 1). It is at this stage, the authors suggest, that the subunits of ClyA come together to form a ring, with the tip of the third helix pointing inwards. And then, in the final step of assembly, the third helices of all 12 subunits tilt into and span the bilayer in an iris-like arrangement to form the pore. Intriguingly, the iris resembles the transmembrane domain of the Wza octamer, which was the first protein that was found[7] to span the outer membrane of a bacterium in helical form.

The suggested route of assembly for ClyA resembles the mechanism elucidated for β-barrel toxins, in which the non-penetrating ring has been termed the pre-pore[8,9] (Fig. 1). In the pre-pore mechanism, concerted, irreversible penetration of the bilayer follows ring formation. The delay before pre-pore formation allows an effective concentration of subunits to gather on the membrane surface before activity is triggered. Moreover, the penetration of subunits with individual β-strands (or α-helices with hydrophilic faces) would be energetically disfavoured, as these structures tend to avoid non-aqueous environments.

The binding and conformational changes required for the assembly of pores may be mediated by receptors[9], although in most cases, including that of ClyA[10], slower spontaneous assembly can occur. For example, the CDCs bind to cholesterol in membranes through a domain that, in the monomer, is distant from the residues that refold into the membrane-spanning strands[11]. Modest contortions of the

protein are then required for bilayer penetration. In other cases, it seems likely that release from the receptor must occur before the pre-pore can proceed to the pore[12]. Interestingly, this problem is solved in the case of ClyA by the extensive reorganization of the polypeptide chain that occurs before pre-pore formation, which brings the membrane-binding domain and the eventual membrane-spanning domain into close proximity (Fig. 1).

Many questions remain about these intriguing proteins, whether they be α-helices or β-barrels. Do the monomers assemble in a random or sequential manner? How does concerted insertion of the transmembrane domains occur? How does the membrane lipid bilayer fit around the pre-pore and the fully assembled pore, and how is the central lipid expelled on pre-pore formation? Can the role of the receptors for these proteins be better defined?

Further structural insights will bolster the application of engineered pores in biotechnology and will help in the design of therapies to either block their action or prevent their assembly. Recently, complement C9 and perforin were shown to be related to the CDCs. Moreover, a vast array of C9 and

perforin-like proteins (perhaps not all of which perforate membranes) are involved in activities as diverse as neuronal migration and parasite exit from host cells[13,14]. Accordingly, further comparative study of pore-forming proteins, including relatives of ClyA, is expected to yield many more fascinating findings[15]. ∎

Hagan Bayley is in the Department of Chemistry, University of Oxford, Oxford OX1 3TA, UK.
e-mail: hagan.bayley@chem.ox.ac.uk

1. Lacovache, I., van der Goot, F. G. & Pernot, L. *Biochim. Biophys. Acta* **1778,** 1611–1623 (2008).
2. Mueller, M., Grauschopf, U., Maier, T., Glockshuber, R. & Ban, N. *Nature* **459,** 726–730 (2009).
3. Song, L. *et al. Science* **274,** 1859–1866 (1996).
4. Heuck, A. P. *et al. Biochemistry* **40,** 9065–9073 (2001).
5. Wallace, A. J. *et al. Cell* **100,** 265–276 (2000).
6. White, J. M., Delos, S. E., Brecher, M. & Schornberg, K. *Crit. Rev. Biochem. Mol. Biol.* **43,** 189–219 (2008).
7. Dong, C. *et al. Nature* **444,** 226–229 (2006).
8. Walker, B. *et al. Chem. Biol.* **2,** 99–105 (1995).
9. Bayley, H., Jayasinghe, L. & Wallace, M. *Nature Struct. Mol. Biol.* **12,** 385–386 (2005).
10 Eifler, N. *et al. EMBO J.* **25,** 2652–2661 (2006).
11. Soltani, C. E., Hotze, E. M., Johnson, A. E. & Tweten, R. K. *Proc. Natl Acad. Sci. USA* **104,** 20226–20231 (2007).
12. Olson, R. & Gouaux, E. *J. Mol. Biol.* **350,** 997–1016 (2005).
13. Rosado, C. J. *et al. Cell. Microbiol.* **10,** 1765–1774 (2008).
14. Kafsack, B. F. *et al. Science* **323,** 530–533 (2009).
15. Anderluh, G. & Lakey, J. H. *Trends Biochem. Sci.* **33,** 482–490 (2008).

**GEOPHYSICS**

# Magnetic ringing of the Earth

Richard Holme

**The study of Earth's core is an exacting task that involves interpreting indirect evidence. An innovative approach to understanding decadal changes in the magnetic field provides a new probe of core dynamics.**

Large earthquakes excite wave motion throughout the Earth, causing it to ring like a bell. The resulting ground motion, measured by seismometers, provides helpful constraints on Earth's internal structure through the study of its 'normal modes' — free oscillations that occur at discrete frequencies. Seismologically measurable oscillations have typical periods of a few minutes. But in work just published in *Geophysical Journal International*, Buffett *et al.*[1] sound the Earth, in particular its core, through oscillations that have periods of decades.

The 'normal modes' studied by Buffett *et al.* are detected, not seismologically, but through the geomagnetic field. This field originates in Earth's molten, predominantly iron, core, being generated by the convection-driven fluid motions of a hydrodynamic dynamo. Global models of this field at Earth's surface have been constructed[2] that cover a period going back 400 years. By assuming that Earth's mantle is at most a very weak electrical conductor, these models reveal decadal-timescale changes in the magnetic field (secular variation) at the core–mantle boundary. The changes in field are then used to probe the physics of the top of the core.

As an analogy, imagine throwing dye into a river; the movement of the dye provides information about the direction and strength of the currents in the river, until gradually, over time, the dye diffuses away. The magnetic field is used as a similar proxy for processes at the top of the core. Further, because the core is an excellent electrical conductor, scaling arguments suggest that the effect of the flow (the advection of the magnetic field) dominates diffusion on decadal timescales. Consequently, secular variation can be used to constrain the structure of the flow at the core's surface[3].

What processes give rise to the flow that causes decadal secular variation? Both magnetic forces and the effect of Earth's rapid rotation impose strong constraints on the dynamics of flow in the core. Analysis of the dynamical equations suggests that core flow can vary on timescales of days (for example, through phenomena controlled by Earth's rotation) or timescales of hundreds of years (a trade-off between magnetic forces, buoyancy forces from variations in density, and rotational effects). On decadal timescales, only one class of motion has been suggested to occur: uniform flow that can
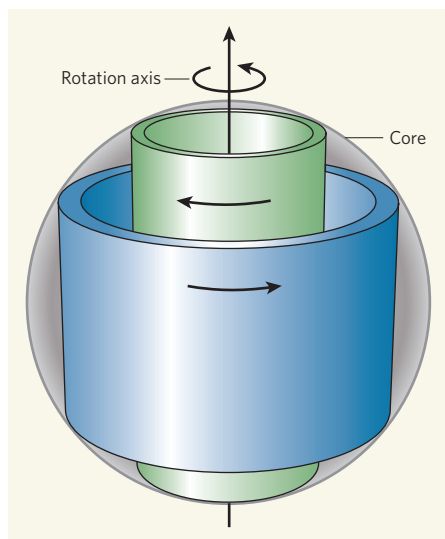
**Figure 1 | Core effects on a timescale of decades.**
Observed changes in Earth's magnetic field on decadal timescales are associated with uniform fluid motion in the form of cylinders lying concentric with respect to Earth's axis of rotation. Modelling of these motions[1] provides a probe of physical conditions deep within Earth's fluid core. (Figure courtesy of S. Zatman.)

be described by the motion of cylinders aligned with the rotation axis (Fig. 1).

These cylindrical motions are special because they are unaffected by rotational (Coriolis) forces — hence their distinct timescale. They also allow processes within the core to be inferred from observations at the surface. Because they are uniform throughout the core, they can be constrained using a model of surface flow calculated from secular variation. Further, they carry changes in angular momentum in the core; when these changes are calculated and compared with changes in the rotation of the solid Earth (the variation in the length of a day), striking agreement is observed[4,5]. The cylinders are linked by the magnetic field that threads them. If they are displaced with respect to each other, a restoring force is generated; the stronger the field in the (cylindrically) radial direction perpendicular to their surface, the stronger the force that binds them together. This force gives rise to oscillations called torsional oscillations[6], which are calculated to have periods of the order of decades.

Modelling the torsional motions promises to constrain not only the radial magnetic field, but also other forces acting on the system: viscous and electromagnetic friction at the surfaces of the outer and inner core, and even the rigidity of the solid inner core due to its gravitational coupling to the mantle. Earlier work[7] attempted to sound the interior of the core by fitting torsional waves to the observed flow. Buffett and colleagues[1] provide a substantial advance on this by expanding the flow on the basis of the normal modes of torsional oscillation[8].

The results provide highly suggestive clues

about various aspects of core dynamics. Buffett *et al.* find evidence of magnetic flux tubes that is consistent with previous inferences from magnetic-field models. They also find indications of a forcing of the oscillations at the cylindrical boundary above and below the inner core, and a profile for at least part of the magnetic field within the core surface (surface observations tell us only about the field at the core surface and above). These observations provide powerful guidance for discriminating between various possible models of the geodynamo.

The methodology developed by Buffett *et al.*[1] is an exciting new tool for investigating the interior of Earth's core. But much remains to be done. The flow is poorly constrained; many possible flows fit the observed geomagnetic-field evolution, so a much wider exploration of the sensitivity of the results to different allowed flows is necessary. The model is not consistent with some observations: the value for the radial field at the core–mantle boundary is lower than observed (itself only a lower bound), and decadal length-of-day variations are poorly fitted — as Buffett *et al.*[1] suggest, further constraining the torsional flows so that they also fit length-of-day data may be useful. Such a constraint might also account for at least part of the excitation of the oscillations; length-of-day data suggest that the core has been subjected to occasional impulsive torques, which are associated with the sharp alterations in the rate of change of secular

variation known as 'geomagnetic jerks'[9].

Fundamental questions remain. It will be crucial to determine what physical mechanism excites the oscillations — is it a natural consequence of the dynamo process; is a more discontinuous process in action (perhaps associated with the slow solidification of the inner core, or core–mantle coupling); or is some external mechanism influencing the dynamics of the core? Such questions can only be asked sensibly if proposed mechanisms can be tested against observations. The work of Buffett *et al.* provides the necessary framework; inaccessible as it is, the interior of the core is gradually giving up some of its secrets. ■

Richard Holme is in the Department of Earth and Ocean Sciences, University of Liverpool, Liverpool L69 3GP, UK.
e-mail: r.t.holme@liv.ac.uk

1. Buffett, B. A., Mound, J. & Jackson, A. *Geophys. J. Int.* **177**, 878–890 (2009).
2. Jackson, A., Jonkers, A. R. T. & Walker, M. R. *Phil. Trans. R. Soc. Lond. A* **358**, 957–990 (2000).
3. Holme, R. in *Treatise on Geophysics.* Vol. 8 *Core Dynamics* (ed. Olson, P.) 107–130 (Elsevier, 2007).
4. Jault, D., Gire, C. & LeMouël, J.-L. *Nature* **333**, 353–356 (1988).
5. Jackson, A., Bloxham, J. & Gubbins, D. in *Dynamics of Earth's Deep Interior* (eds LeMouël, J.-L., Smylie, D. E. & Herring, T. A.) 97–107 (Am. Geophys. Union, 1993).
6. Braginsky, S. I. *Geomag. Aeron., Engl. Transl.* **10**, 1–8 (1970).
7. Zatman, S. & Bloxham, J. *Nature* **388**, 760–763 (1997).
8. Mound, J. E. & Buffett, B. A. *J. Geophys. Res.* **110**, B08103 (2005).
9. Holme, R. & de Viron, O. *Geophys. J. Int.* **160**, 435–439 (2005).

**QUANTUM MECHANICS**

# Entanglement goes mechanical

Rainer Blatt

**A neat experiment shows that the mechanical vibration of two ion pairs separated by a few hundred micrometres is entangled — their motions are intrinsically and inseparably connected in a quantum way.**

For almost a century, physicists have had in hand 'the' theoretical framework of the known world — quantum mechanics. But whereas the world clearly comprises large complex systems, quantum mechanics is usually associated with the microworld of atoms and elementary particles, and is hardly ever considered as an underlying feature in our daily life. This is even more pronounced for some of the seemingly weird predictions of quantum mechanics, such as entanglement, which asserts that the quantum state of physically separated objects is mutually and inextricably connected.

On page 683 of this issue, Jost and colleagues[1] demonstrate quantum entanglement of two spatially separated mechanical oscillators. Although the quantum nature of mechanical oscillators has been known and observed for a long time, the entanglement of their oscillating motions has not, and its

demonstration adds a valuable tool to the toolbox of quantum-state engineering.

Entanglement was predicted back in the 1930s (ref. 2), and has been explored since the 1980s as one of the odd features of quantum mechanics. However, the routine production of entangled states of atoms and elementary particles became possible only during the past 10–15 years[3], which have witnessed the use of such states in quantum-information science and quantum metrology as well as in the investigation of some of the most fundamental problems in quantum physics.

In their experiment, Jost *et al.*[1] manipulate two pairs of ions, each pair consisting of two different positively ionized atom species. After confining the ions in an arrangement of electrodes using high-frequency voltages — an electromagnetic trap known as a Paul trap — the authors use laser radiation and slowly

varying trapping voltages to hold and position the ions on a chip-like electrode structure. Because the ions are positively charged, they repel each other via Coulomb repulsion. Laser irradiation enables the ions' motion to be frozen and manipulates their internal state as well as their vibrational motion. Application of additional manipulation voltages separates and arranges the ions into two pairs, which are then kept trapped and oscillating in two distant and independent locations. Each ion pair behaves as a mechanical oscillator, and can be thought of as a miniature pendulum of two atoms that vibrate in parallel or against one another, or — as depicted in Figure 1 of the paper[1] on page 683 — as two masses connected by a spring a few micrometres long.

Using their new technique[1] for manipulating ion pairs and their vibrational motions, which involves quite sophisticated sequences of laser pulses and pulsed voltages, Jost and colleagues were able to entangle the mechanical oscillation of the two vibrating subsystems. These are a few hundred micrometres apart and thus can definitely be viewed as independent mechanical systems at different positions. But being entangled, the two seemingly unconnected vibrating strings move as if magically linked together — that is, their motions can no longer be described independently of one another, and must instead be explained by a joint quantum state. The micromechanical motion of the subsystems is an entangled mechanical vibration.

To create micromechanical entanglement, Jost et al. have developed a number of atomic-manipulation and detection techniques that will ultimately pave the way for chip technology using atoms as quantum information storage sites. Moreover, the experiment required the realization of new protocols and procedures that will be extremely useful for quantum-metrology and quantum-information applications. One such procedure is the repeated transfer of entanglement from an atom's internal state (the spin state) to the micromechanical motion of the ion pairs.

Whereas Jost and colleagues and other 'ion trappers' around the world have been able to entangle atoms previously[3], with a record number of eight atoms[4], the new study[1] is the first to achieve entanglement of the mechanical motion of pairs of ionic atoms. Apart from adding another toy to the quantum mechanic's playground, this is an important tool for further developments in quantum-state engineering. For example, such entanglement could be used to distribute quantum information on an atom (ion) chip or, more generally, to construct mechanical sensors. What's more,

it provides the basic ingredient for generating optomechanical interfaces such as atom–micromirror systems[5].

Mechanical entanglement, together with the authors' techniques for manipulating and controlling two atom species, will lead to a new generation of optomechanical experiments, which will allow the introduction of 'quantumness' into the micro- and macroscopic mechanical world. Eventually, such techniques will constitute the basic building blocks — if not the genetic code — for the realization of a 'quantum pet', Schrödinger's cat[2], which is a quantum superposition state of a macroscopic object. With the availability of such macroscopic quantum systems, the bizarre quantum world will one day become more tangible to us. ∎

Rainer Blatt is at the Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Innsbruck, and the Institut für Experimentalphysik, Universität Innsbruck, Innsbruck A-6020, Austria.
e-mail: rainer.blatt@oeaw.ac.at

1. Jost, J. D. et al. Nature 459, 683–685 (2009).
2. Schrödinger, E. Naturwissenschaften 23, 807–812; 823–828; 844–849 (1935).
3. Nature Insight on Quantum Coherence Nature 453, 1003–1049 (2008).
4. Häffner, H. et al. Nature 438, 643–646 (2005).
5. Kippenberg, T. & Vahala, K. Science 321, 1172–1176 (2008).

---

## DEVELOPMENTAL BIOLOGY

# The early heart remodelled

Fu-Sen Liang and Gerald R. Crabtree

**What factors direct the formation of heart muscle in the developing embryo? Unexpectedly, a chromatin-remodelling protein complex turns out to be a crucial determinant of cardiac-cell fate.**

The chances are that many of us will develop heart disease at some point in our lives. So understanding how heart muscle, heart valves and other heart structures regenerate is particularly important. The work published by Takeuchi and Bruneau in this issue (page 708)[1] takes us a step closer to achieving this goal, and also provides us with intriguing insights into the role of chromatin-remodelling complexes in the developing heart.

ATP-dependent chromatin-remodelling complexes are multi-protein machines that use the energy of ATP hydrolysis to modify the structure of chromatin (the complex of DNA and histone proteins), thereby activating or repressing genes. The vertebrate chromatin-remodelling BAF complexes resemble the yeast chromatin-remodelling complexes SWI/SNF, RSC and SWR. However, whereas each yeast complex is probably made up of an invariant set of protein subunits, in BAF complexes, combinatorial assembly of subunits produces distinct functional complexes. For example,

mouse BAF complexes consist of 11 subunits; one of these is the Baf60 subunit, which is encoded by three genes, *Baf60a*, *Baf60b* and *Baf60c* (ref. 2). BAF complexes usually repress their target genes, whereas yeast SWI/SNF activates all known targets[3,4], implying that there are mechanistic as well as structural differences between BAF complexes and SWI/SNF. So far, how these complexes regulate gene expression is a matter of speculation.

Takeuchi and Bruneau[1] focused on BAF complexes because these proteins have been implicated in cardiac development — a screen conducted in early mouse embryos found that *Baf60c* was expressed selectively in tissues that give rise to the heart[5]. Notably, when *Baf60c* was depleted experimentally, heart formation was defective[5]. Takeuchi and Bruneau reasoned that, as *Baf60c* is selectively expressed in mouse embryonic cardiac tissue, it might contribute to heart formation if injected into the embryo outside the heart developmental fields. Indeed, when the authors inject *Baf60c* together with

the cardiogenic transcription factor *Gata4* outside the heart fields, these tissues express heart-specific genes[1]. Baf60c and Gata4 also induce expression of another transcription factor, Nkx2-5 (ref. 1). Thus, it seems that Gata4 cooperates with Baf60c to set up a gene-expression program that drives heart development. And from what we know about BAF complexes, it seems likely that the transfected Baf60c assembles with the other ten subunits to form a cardiogenic BAF complex (Fig. 1). Remarkably, addition of yet another transcription factor, Tbx5, transforms non-cardiac embryonic cells into beating heart muscle[1].

So how do Gata4 and Baf60c cooperate to induce the expression of cardiac-specific genes? On the basis of results obtained from chromatin immunoprecipitation studies (a technique that detects the binding of proteins to DNA using antibodies), the authors propose that Gata4 binds directly to Baf60c (Fig. 1). They also find that Baf60c doesn't function by interacting with the general transcriptional apparatus. Instead, it is required for Gata4 binding to DNA (Fig. 1), consistent with its apparently instructive role in this model system. Interestingly, this is the opposite to the situation in yeast, in which the SWI5 transcription factor first binds to DNA and then recruits the SWI/SNF complex to its target promoter[6]. BAF complexes contain domains that can interact with modified histones and DNA; hence, in the absence of a transcription-factor 'guide', they might find their targets by multivalent
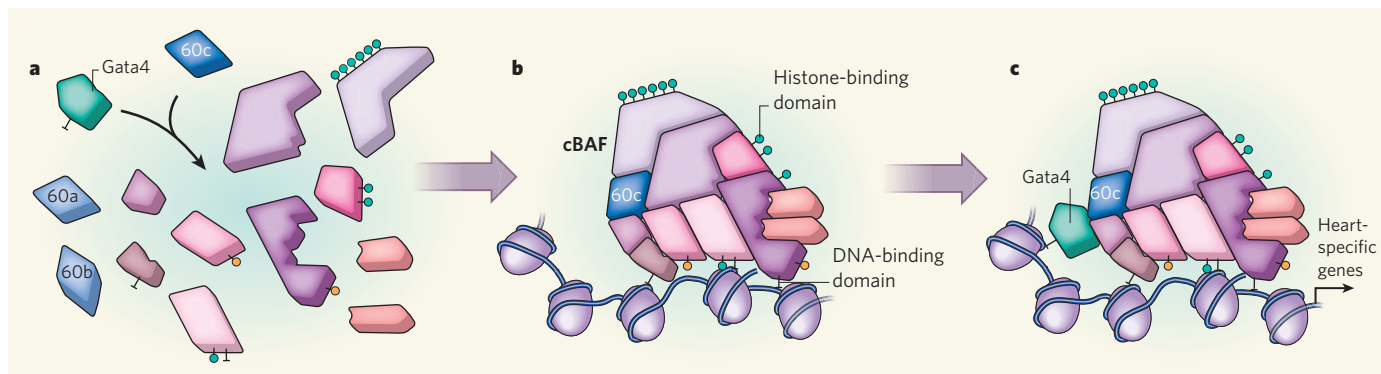
**Figure 1 | A chromatin-remodelling complex instructs heart development.** Takeuchi and Bruneau[1] induced the expression of the cardiac-specific Baf60c subunit and the transcription factor Gata4 in mouse embryonic tissues outside the heart developmental fields. **a**, Transfected Baf60c probably assembles with the other ten BAF subunits to form a cardiogenic BAF complex (cBAF). Baf60a and Baf60b are not incorporated into cBAF. **b**, cBAF seems to bind to target loci, perhaps by multivalent recognition through histone- and DNA-binding domains. **c**, cBAF recruits Gata4 through a presumed interaction between Baf60c and Gata4, which results in the expression of heart-specific genes. The arrangement of subunits in the BAF complex is not yet established.

recognition of a specific genetic locus[7].

Apparent cooperation between chromatin-remodelling complexes and tissue-specific transcription factors is not unique to the developing heart. A similar process occurs in embryonic stem cells, in which a variant of the BAF complex that is apparently found only in these cells (esBAF) interacts with Oct4 and Sox2 transcription factors[8]. esBAF complexes are essential for embryonic-stem-cell formation and also for formation of the inner cell mass — the part of the embryo that is pluripotent, giving rise to all cell types. The distinctive subunit composition of esBAF[9] seems to produce its biological specificity, similar to the way that letters are combined into words with specific meanings. Indeed, misspelling the esBAF 'word', by incorporating incorrect subunits, results in a loss of pluripotency.

There are also specialized chromatin-remodelling complexes in neural progenitors (npBAF) and neurons (nBAF), and misspelling the 'word' in these cells, for instance by genetic ablation of one subunit and replacement with another, leads to lethal defects in the developing nervous system[10]. In each neural complex there seems to be specific interactions between the chromatin-remodelling complex and the specific cellular transcription factors.

The instructive role of BAF complexes in cellular differentiation is unexpected, as previous biochemical studies had suggested that the complexes have a nonspecific role in transcription, either functioning as part of the general transcriptional machinery or directing the binding of TBP (a transcription initiation factor) to promoters. But even in yeast, the association of SWI/SNF with general transcription factors or with transcription-associated RNA polymerase has not been consistently reproduced. Takeuchi and Bruneau's findings are, however, consistent with the observation that two subunits of the npBAF complex, Baf45a and Baf53a, can direct the division of neural stem cells in places in the developing nervous system where these cells don't usually proliferate[11].

The emerging model of the genetic circuitry

that drives heart development[1,12] (Fig. 2) now suggests that Baf60c may be near the top of the hierarchy. This instructive role is not due to its tissue-specific expression, because it is also found in npBAF complexes[11,13]. Perhaps Baf60c has some general role in the control of multipotent cells, by modifying chromatin structure and allowing genes to remain accessible for later activation. Alternatively, it could achieve its specificity by combining with other BAF subunits that have restricted expression, generating an assembly of subunits found only



**Figure 2 | The genetic circuitry regulating heart development.** In mammals, extracellular signals act directly or indirectly to produce localized expression of the BAF complex subunit Baf60c and the transcription factor Gata4 in the area of the embryo that gives rise to the heart. Baf60c and Gata4 work together with another cardiogenic transcription factor, Tbx5, to activate genes encoding other regulators of heart development — Mef2, Nkx2-5, Hand and Isl1. These directly activate certain muscle-specific genes. Additional steps, requiring Tbx5, organize synchronous activity of cardiac muscle cells, giving rise to beating heart muscle. *(Modified from ref. 12.)*

in the regions of the embryo that give rise to the heart.

The details of assembly of the cardiogenic BAF complex (let's call it cBAF) suggest that Takeuchi and Bruneau's strategy[1] for generating heart tissue might work even better if they could reduce expression of the competing non-cardiogenic Baf60 subunits (Baf60a and Baf60b), which might have instructive roles for other tissues. Additional biochemical studies will be required to determine the subunit composition of this complex or family of complexes — a challenging task, given the tiny number of cells that are involved in initially specifying the heart.

Clearly, the next step will be to determine if skin or blood cells can be transformed directly into beating heart muscle using Baf60c and its co-regulators. An alternative approach might be to use induced pluripotent stem cells — cells from adult tissue that have been reprogrammed into stem cells — and to exploit our clearer knowledge of the genetic network involved in cardiac development[1,12] to control heart formation. These strategies might allow direct regeneration of defective or diseased heart muscle using a patient's own tissue, without the need for immunosuppression. ■

Fu-Sen Liang and Gerald R. Crabtree are in the Departments of Pathology and Developmental Biology, Stanford University Medical School, Stanford, California 94305, USA.
e-mail: crabtree@stanford.edu

1. Takeuchi, J. K. & Bruneau, B. G. *Nature* **459,** 708–711 (2009).
2. Wang, W. *et al. Genes Dev.* **10,** 2117–2130 (1996).
3. Biggar, S. R. & Crabtree, G. R. *EMBO J.* **18,** 2254–2264 (1999).
4. Martens, J. A., Wu, P. Y. & Winston, F. *Genes Dev.* **19,** 2695–2704 (2005).
5. Lickert, H. *et al. Nature* **432,** 107–112 (2004).
6. Cosma, M. P. *et al. Cell* **97,** 299–311 (1999).
7. Wu, J. I., Lessard, J. & Crabtree, G. R. *Cell* **136,** 200–206 (2009).
8. Ho, L. *et al. Proc. Natl Acad. Sci. USA* **106,** 5187–5191 (2009).
9. Ho, L. *et al. Proc. Natl Acad. Sci. USA* **106,** 5181–5186 (2009).
10. Wu, J. *et al. Neuron* **56,** 94–108 (2007).
11. Lessard, J. *et al. Neuron* **55,** 201–215 (2007).
12. Olson, E. N. *Science* **313,** 1922–1927 (2006).
13. Lamba, D. A., Hayes, S., Karl, M. O. & Reh, T. *Dev. Dyn.* **237,** 3016–3023 (2008).

# Attrition of memory CD8 T cells

**Arising from: V. Vezys et al. Nature 457, 196–199 (2009)**

An important role for the immune system is to maintain protective immunological memory to a wide variety of pathogens encountered over one's lifetime, while still leaving the host able to respond to newly encountered pathogens. Vezys et al.[1] make the interesting observation that it is possible to repeatedly immunize mice in ways that allow for development of high numbers of memory CD8 T cells without depleting pre-existing memory cells specific for other pathogens. This study, which offers promise in developing potent vaccination schemes, is seemingly at odds with work published by us in the 1990s showing a loss in CD8 memory cells after a series of infections[2,3]. In their reply, Vezys et al. mention that we may have misinterpreted our data because we reported the putative loss of memory T cells as per cent rather than total number, but here we represent the data in those studies as total cell number[2–4]. We show here in Fig. 1 that a series of infections can indeed reduce the total number of memory cells, indicating that vaccination strategies need to consider this issue.



Figure 1a portrays limiting dilution assay data, a measure mostly of central memory cells, with 3 to 11 determinations per group, and Fig. 1b shows determinations with intracellular cytokine and MHC-dimer assays, with 6 determinations per group. They show that losses in memory CD8 cells became statistically significant after a series of viral infections. Virus-induced losses in total body memory have also been reported by in vivo cytotoxicity assays, which monitor the ability of T cells to kill antigen-expressing targets in the host[5]. Others have reported bacteria-induced losses of T-cell-dependent immunity to tumours[6]. These findings differ with the conclusions of Vezys et al.[1], who examined loss of memory cells after infections with only a single pathogen or after a series of immunizations with vectors encoding the same antigen in a prime–boost immunization scheme.

The relevance of that point rests on the observation that virus-induced type 1 interferon (IFN) mediates high levels of apoptosis of memory T cells early during infections[7–10]. We previously addressed the question of whether the long-term loss of memory cells was a function of this early apoptosis (active attrition) or due to a competition model for niches in the immune system (passive attrition), and concluded that most of the loss in memory cells could be accounted for by active attrition rather than competition[5]. We question whether the prime–boost immunization scheme of Vezys et al.[1] would elicit the apoptosis-inducing levels of type 1 IFN that a series of different viruses would, because the homologous immunity elicited to their immunizing antigen inserts would inhibit replication of the vectors. Furthermore, recombinant vaccinia virus like that used in this study is a poor IFN inducer. A recent paper[11] reports a substantial type 2 IFN (IFN-γ)-dependent permanent loss of memory CD8 T cells in mice challenged with intracellular bacteria that were strong IFN-γ (but not type 1 IFN) inducers. There it was also concluded[11] that the loss of memory cells was due to an active process and not due to competition. This indicates that more than one class of IFN can cause memory T-cell attrition.

Vezys et al.[1] state that an argument for memory CD8 T-cell stability is a human study showing a half-life of 8–16 years for vaccinia-virus-specific CD8 T cells[12]. However, that same study could not detect vaccinia-virus-specific CD8 T cells in 50% of the subjects vaccinated over 20 years earlier. We contend that those data are more consistent with a loss in human memory T cells than of memory cell stability.

We therefore argue that the study of Vezys et al.[1] does not refute earlier work showing that a series of infections depletes splenic memory CD8 cells specific to previously encountered pathogens. It does, however, nicely show how one can avoid such depletion by using a prime–boost immunization scheme with vectors that are not strong IFN inducers.

**Raymond M. Welsh**[1] **& Liisa K. Selin**[1]

[1]Department of Pathology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.
e-mail: raymond.welsh@umassmed.edu

**Figure 1 | Depletion of memory CD8 T cells by heterologous viral infections.** Total numbers of lymphocytic choriomeningitis virus (LCMV)-epitope-specific CD8 T cells per spleen in mice immune to LCMV and subsequently infected with Pichinde virus (PV), vaccinia virus (VV), murine cytomegalovirus (MCMV) or vesicular stomatitis virus (VSV). Data are reworked from refs 3 and 4. These are controlled experiments in which all measurements were taken at the same time after LCMV infection. **a,** Attrition of LCMV-specific memory cells as measured by limiting dilution assays after sequential virus infections. **b,** Attrition of LCMV-specific memory cells as measured by MHC-dimer or intracellular IFN-γ assays after sequential virus infections. Combined frequencies of T cells specific to LCMV-encoded epitopes NP396, GP33/34 and GP276 (GP, glycoprotein; NP, nucleoprotein).

1. Vezys, V. et al. Memory CD8 T-cell compartment grows in size with immunological experience. Nature **457**, 196–199 (2009).
2. Selin, L. K., Vergilis, K., Welsh, R. M. & Nahill, S. R. Reduction of otherwise remarkably stable virus-specific cytotoxic T lymphocyte memory by heterologous viral infections. J. Exp. Med. **183**, 2489–2499 (1996).
3. Selin, L. K. et al. Attrition of T cell memory:selective loss of lymphocytic choriomeningitis virus (LCMV) epitope-specific memory CD8 T cells following infections with heterologous viruses. Immunity **11**, 733–742 (1999).
4. Varga, S. M., Selin, L. K. & Welsh, R. M. Independent regulation of lymphocytic choriomeningitis virus-specific T cell memory pools: relative stability of CD4 memory under conditions of CD8 memory T cell loss. J. Immunol. **166**, 1554–1561 (2001).
5. Kim, S. K. & Welsh, R. M. Comprehensive early and lasting loss of memory CD8 T cells and functional memory during acute and persistent viral infections. J. Immunol. **172**, 3139–3150 (2004).

6. Smith, D. K. et al. Cross-reactive antigen is required to prevent erosion of established T cell memory and tumor immunity: a heterologous bacterial model of attrition. *J. Immunol.* **169**, 1197–1206 (2002).

7. McNally, J. M. et al. Attrition of bystander CD8 T cells during virus-induced T cell and interferon responses. *J. Virol.* **75**, 5965–5976 (2001).

8. Jiang, J., Lau, L. L. & Shen, H. Selective depletion of nonspecific T cells during the early stage of immune responses to infection. *J. Immunol.* **171**, 4352–4358 (2003).

9. Bahl, K. et al. Interferon-induced attrition of CD8 T cells in the presence or absence of cognate antigen during the early stages of viral infections. *J. Immunol.* **176**, 4284–4295 (2006).

10. Jiang, J., Gross, D., Nogusa, S., Elbaum, P. & Murasko, D. M. Depletion of T cells by type I interferon: differences between young and aged mice. *J. Immunol.* **175**, 1820–1826 (2005).

11. Dudani, R., Murali-Krishna, K., Krishnan, L. & Sad, S. IFN-γ induces the erosion of preexisting CD8 T cell memory during infection with a heterologous intracellular bacterium. *J. Immunol.* **181**, 1700–1709 (2008).

12. Hammarlund, E. et al. Duration of antiviral immunity after smallpox vaccination. *Nature Med.* **9**, 1131–1137 (2003).

**doi:**10.1038/nature08091

# Vezys et al. *reply*

**Replying to:** R. M. Welsh and L. K. Selin *Nature* **459**, doi:10.1038/nature08091 (2009)

We reported that it is possible to increase the total number of memory CD8 T cells within an organism, and to establish preternatural numbers of vaccine-specific effector memory CD8 T cells while preserving naive CD8 T cells and most pre-existing memory CD8 T cells specific for a previously encountered infection[1]. These findings raise new questions regarding the regulation and limits of generating CD8 T cell immunity. Our discussion highlighted three points related to the issue of attrition. First, that it is possible to over-estimate perceived attrition by only examining percentages (see Fig. 1 of ref. 1)[2]. Second, our vaccine regimen resulted predominantly in the generation of effector memory CD8 T cells located outside of lymph nodes. It remains possible that the number of lymph node central memory T cells remains tightly regulated. Third, we noted that our data did not refute that attrition could happen under a variety of circumstances. However, our data demonstrate that attrition is not an axiomatic property of immunization, mandated by stringent regulation of the size of the total memory CD8 T-cell compartment. Indeed, we saw no evidence of attrition after single infections with a virus (vaccinia), an intracellular bacteria (*Listeria monocytogenes*) and a parasite that induces massive splenomegaly (*Plasmodium yoelii*), and observed comparatively little attrition after a heterologous prime–boost regimen involving successive immunization with three viruses[1].

The issue at hand is whether our results represent rare exceptions that could be exploited for vaccination, or whether preservation of immunological memory is a normal feature of the mammalian immune system. Welsh and Selin[2] argue that memory T-cell maintenance is regulated by the quantity of type 1 IFN produced in response to subsequent infections, and that most infections invariably lead to profound loss of pre-existing memory CD8 T cells. Data supporting this claim demonstrate that up to 25–90% of pre-existing memory CD8 T cells are lost after single infections, and that attrition accumulates further with each infection[3,4]. Surprisingly, infection-specific naive and memory T cells were not spared[5]. The current model of Welsh and Selin[2] predicts that immunological memory must be short-lived in the face of occasional infections. Consequently, they question the evidence that CD8 T-cell memory specific for vaccinia has a long half-life in humans[6]. Memory CD8 T cells have also been examined after a single exposure to both measles and hantavirus, and robust populations could be detected >10 years later[7,8]. Protective immunity against measles virus (for which T-cell memory may be important) persists >65 years (ref. 9). Memory B cells persist indefinitely after smallpox immunization, and the half-life of serum antibody, which must be continuously produced by differentiated antigen-experienced B cells, ranges from 92 to 3,014 years after exposure to measles, mumps, rubella or vaccinia virus[10,11]. These data indicate that immunological memory does not necessarily undergo rapid erosion, although further studies regarding the longevity of CD8 T memory are needed.

The concept advanced by our paper is that attrition is not an axiomatic property of immunization, and that the memory CD8 T-cell compartment is capable of expansion. In support of these findings, it was recently reported that primary human cytomegalovirus infection resulted in a long-lived increase in the total number of antigen-experienced CD8 T cells, and induced a reduction in the frequency, but not number, of pre-existing memory CD8 T cells specific for influenza or Epstein–Barr viruses in blood[12]. It should be stressed that immunity to certain agents (for example, after vaccination with non-replicating agents) may be intrinsically prone to attrition, and certain infections may significantly erode immunological memory. These may include pathogens that infect lymphocytes, destroy lymphoid tissue or are associated with other pathologies, as indicated by Welsh and Selin[2]. Although our results raise the bar for what levels of CD8 T-cell memory might be achievable through vaccination, Welsh and Selin[2] highlight the important fact that each vaccine vector should be evaluated empirically for safety and the preservation of pre-existing immunological memory.

**Vaiva Vezys**[1,2], **Andrew Yates**[3], **Kerry A. Casey**[1], **Gibson Lanier**[2], **Rafi Ahmed**[2], **Rustom Antia**[3] & **David Masopust**[1,2]

[1]Department of Microbiology and Center for Immunology, University of Minnesota, Minneapolis, Minnesota 55455, USA.
e-mail: masopust@umn.edu
[2]Emory Vaccine Center, Emory University School of Medicine, Atlanta, Georgia 30322, USA.
[3]Department of Biology, Emory University, Atlanta, Georgia 30322, USA.

1. Vezys, V. et al. Memory CD8 T-cell compartment grows in size with immunological experience. *Nature* **457**, 196–199 (2009).

2. Welsh, R. M. & Selin, L. K. Attrition of memory CD8 T cells. *Nature* **459**, doi:10.1038/nature08091 (2009).

3. Dudani, R., Murali-Krishna, K., Krishnan, L. & Sad, S. IFN-γ induces the erosion of preexisting CD8 T cell memory during infection with a heterologous intracellular bacterium. *J. Immunol.* **181**, 1700–1709 (2008).

4. Selin, L. K. et al. Attrition of T cell memory: selective loss of LCMV epitope-specific memory CD8 T cells following infections with heterologous viruses. *Immunity* **11**, 733–742 (1999).

5. Bahl, K. et al. IFN-induced attrition of CD8 T cells in the presence or absence of cognate antigen during the early stages of viral infections. *J. Immunol.* **176**, 4284–4295 (2006).

6. Hammarlund, E. et al. Duration of antiviral immunity after smallpox vaccination. *Nature Med.* **9**, 1131–1137 (2003).

7. Van Epps, H. L. et al. Long-lived memory T lymphocyte responses after hantavirus infection. *J. Exp. Med.* **196**, 579–588 (2002).

8. Nanan, R., Rauch, A., Kampgen, E., Niewiesk, S. & Kreth, H. W. A novel sensitive approach for frequency analysis of measles virus-specific memory T-lymphocytes in healthy adults with a childhood history of natural measles. *J. Gen. Virol.* **81**, 1313–1319 (2000).

9. Panum, P. L. *Observations Made During the Epidemic of Measles on the Faroe Islands in the Year 1846* (Delta Omega Society, 1940).

10. Crotty, S. et al. Cutting edge: long-term B cell memory in humans after smallpox vaccination. *J. Immunol.* **171**, 4969–4973 (2003).

11. Amanna, I. J., Carlson, N. E. & Slifka, M. K. Duration of humoral immunity to common viral and vaccine antigens. *N. Engl. J. Med.* **357**, 1903–1915 (2007).

12. van Leeuwen, E. M. et al. Differential usage of cellular niches by cytomegalovirus versus EBV- and influenza virus-specific CD8+ T cells. *J. Immunol.* **177**, 4998–5005 (2006).

**doi:**10.1038/nature08092

# ARTICLES

# Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes

Geraldine Butler[1], Matthew D. Rasmussen[2], Michael F. Lin[2,3], Manuel A. S. Santos[4], Sharadha Sakthikumar[3], Carol A. Munro[5], Esther Rheinbay[2,6], Manfred Grabherr[3], Anja Forche[7], Jennifer L. Reedy[8], Ino Agrafioti[9], Martha B. Arnaud[10], Steven Bates[11], Alistair J. P. Brown[5], Sascha Brunke[12], Maria C. Costanzo[10], David A. Fitzpatrick[1], Piet W. J. de Groot[13], David Harris[14], Lois L. Hoyer[15], Bernhard Hube[12], Frans M. Klis[13], Chinnappa Kodira[3]†, Nicola Lennard[14], Mary E. Logue[1], Ronny Martin[12], Aaron M. Neiman[16], Elissavet Nikolaou[5], Michael A. Quail[14], Janet Quinn[17], Maria C. Santos[4], Florian F. Schmitzberger[10], Gavin Sherlock[10], Prachi Shah[10], Kevin A. T. Silverstein[18], Marek S. Skrzypek[10], David Soll[19], Rodney Staggs[18], Ian Stansfield[5], Michael P. H. Stumpf[9], Peter E. Sudbery[20], Thyagarajan Srikantha[19], Qiandong Zeng[3], Judith Berman[7], Matthew Berriman[14], Joseph Heitman[8], Neil A. R. Gow[5], Michael C. Lorenz[21], Bruce W. Birren[3], Manolis Kellis[2,3]* & Christina A. Cuomo[3]*

***Candida* species are the most common cause of opportunistic fungal infection worldwide. Here we report the genome sequences of six *Candida* species and compare these and related pathogens and non-pathogens. There are significant expansions of cell wall, secreted and transporter gene families in pathogenic species, suggesting adaptations associated with virulence. Large genomic tracts are homozygous in three diploid species, possibly resulting from recent recombination events. Surprisingly, key components of the mating and meiosis pathways are missing from several species. These include major differences at the mating-type loci (*MTL*); *Lodderomyces elongisporus* lacks *MTL*, and components of the a1/α2 cell identity determinant were lost in other species, raising questions about how mating and cell types are controlled. Analysis of the CUG leucine-to-serine genetic-code change reveals that 99% of ancestral CUG codons were erased and new ones arose elsewhere. Lastly, we revise the *Candida albicans* gene catalogue, identifying many new genes.**

Four *Candida* species, *C. albicans*, *C. glabrata*, *C. tropicalis* and *C. parapsilosis*, together account for ~95% of identifiable *Candida* infections[1]. Although *C. albicans* is still the most common causative agent, its incidence is declining and the frequency of other species is increasing. Of these, *C. parapsilosis* is a particular problem in neonates, transplant recipients and patients receiving parenteral nutrition; *C. tropicalis* is more commonly associated with neutropenia and malignancy. Other *Candida* species, including *C. krusei*, *C. lusitaniae* and *C. guilliermondii*, account for <5% of invasive candidiasis. Almost all *Candida* species, with the exceptions of *C. glabrata* and *C. krusei*, belong in a single *Candida* clade (Fig. 1) characterized by the unique translation of CUG codons as serine rather than leucine[2]. Within this, haploid and diploid species occupy two separate subclades (Fig. 1).

To determine the genetic features underlying their diversity of biology and pathogenesis, we sequenced six genomes from the *Candida* clade (Fig. 1). These include a second sequenced isolate of

*C. albicans* (WO-1) characterized for white–opaque switching, a phenotypic change that correlates with host specificity and mating[3,4]. We also sequenced the major pathogens *C. tropicalis* and *C. parapsilosis*; *L. elongisporus*, a close relative of *C. parapsilosis* recently identified as a cause of bloodstream infection[5]; and two haploid emerging pathogens, *C. guilliermondii* and *C. lusitaniae*. We compared these with the previously sequenced *C. albicans* strain (SC5314)[6–8], *Debaryomyces hansenii*[9], a marine yeast rarely associated with disease, and nine species from the related *Saccharomyces* clade (Fig. 1). These species span a wide evolutionary range and show large phenotypic differences in pathogenicity and mating, allowing us to study the genomic basis for these traits.

## Genome sequence and comparative annotation

We found enormous variation in genome size and composition between the *Candida* genomes sequenced (Table 1). Each genome assembly displayed high continuity, ranging from nine to 27 scaffolds

**Figure 1 | Phylogeny of sequenced *Candida* and *Saccharomyces* clade species.** Tree topology and branch lengths were inferred with MRBAYES (Supplementary Information, section S5). Posterior probabilities are indicated for each branch. The asterisk marks a branch that was constrained on the basis of syntenic conservation[40].

(Supplementary Table 1). Scaffold number and size largely match pulsed-field gel electrophoresis estimates for all genomes, and telomeric repeat arrays are linked to the ends of nearly all large scaffolds (Supplementary Information, section S2). Genome size ranges from 10.6 to 15.5 megabases (Mb), a difference of nearly 50%, with haploid species having smaller genomes. GC content ranges from 33% to 45% (Table 1). Transposable elements and other repetitive sequences vary in number and type between assemblies (Supplementary Information, section S6). Regions similar to the major repeat sequence (MRS) of *C. albicans* were found only in *C. tropicalis*, suggesting that MRS-associated recombination could contribute to the observed karyotypic variation among *C. tropicalis* strains[10].

Despite the genome size and phenotypic variation among the species, the predicted numbers of protein-coding genes are very similar, ranging from 5,733 to 6,318 (Table 1). Even the small differences in gene number are not correlated with genome size; the smallest genome, *C. guilliermondii*, has more genes than the largest genome, *L. elongisporus*. Instead, genome size differences are explained by an approximately threefold variation in intergenic spacing (Table 1). Large syntenic blocks of conserved gene order were detected among the four diploid species and between two of the haploid species, *C. guilliermondii* and *D. hansenii* (Supplementary Fig. 5). Although synteny blocks have been shuffled by local inversions and rearrangements, these have been primarily intrachromosomal as chromosome boundaries have been largely preserved across the diploid genomes (Supplementary Fig. 5).

Given the high conservation of protein-coding genes across the *Candida* clade, we used multiple alignments of the related genomes to revise the annotation of *C. albicans*. We identified 91 new or updated genes, of which 80% are specific to the *Candida* clade

(Supplementary Information, section S4). We also corrected existing annotations in *C. albicans*, revealing 222 dubious genes, and also identified 190 probable frame shifts and 36 nonsense sequencing errors in otherwise well-conserved genes (Supplementary Information, section S4). In each case, manual curation confirmed ~80% of these predictions.

### Polymorphism in diploid genomes

To gain insights into the recent history of *C. albicans*, we compared the two diploid strains, SC5314 and WO-1, which belong to different population subgroups[11]. Variation in the karyotype of these strains is primarily due to translocations at MRS sequences (ref. 12 and Supplementary Fig. 1). The two assemblies are largely co-linear with 12 inversions of 5–94 kilobases between them, except that in WO-1 some non-homologous chromosomes have recombined at the MRS (Supplementary Information, section S8). We found similar rates of single nucleotide polymorphisms (SNPs) within each strain (one SNP per 330–390 bases), and twice this rate between them, suggesting relatively recent divergence. Polymorphism rates in the other diploids range from one SNP per 222 bases in *L. elongisporus* and one SNP per 576 bases in *C. tropicalis* to a particularly low one SNP per 15,553 bases in *C. parapsilosis*, more than 70-fold lower than in the closely related *L. elongisporus*.

Notable regions of extended homozygosity are found in three of the four diploid genomes, which may reflect break-induced replication or recent passage through a parasexual or sexual cycle. *Candida albicans*, *C. tropicalis* and *L. elongisporus* each shows large chromosomal regions devoid of SNPs, extending up to ~1.2 Mb (Fig. 2 and Supplementary Figs 6–8). In contrast, the few SNPs in *C. parapsilosis* are randomly distributed across the genome (Supplementary Fig. 9a). A total of 4.3 Mb (30%) of the WO-1 assembly is homozygous for SNPs, approximately twice that found in SC5314 (Supplementary Information, section 7, and refs 7, 13, 14). There is at least one homozygous region per chromosome, none of which spans the predicted centromeres and only one of which starts at a MRS (Fig. 2). Whereas nearly all homogeneous regions are present at diploid levels and are therefore homozygous, WO-1 has lost one copy of a >300-kilobase region on chromosome 3 comprising nearly 200 genes (Supplementary Fig. 10). The pressure to maintain this region as homozygous in both strains is apparently high, as it is diploid but homogenous in SC5314.

### Usage and evolution of CUG codons

All *Candida* clade species translate CUG codons as serine instead of leucine[15]. This genetic-code change altered the decoding rules of CUN codons in the *Candida* clade: whereas *Saccharomyces cerevisiae* uses two transfer RNAs, each of which translates two codons, *Candida* species use a dedicated $tRNA_{CAG}^{Ser}$ for CUG codons and a single $tRNA_{IAG}^{Leu}$ for CUA, CUC, and CUU codons, as inosine can base-pair with A, C and U (Fig. 3). This alteration in decoding rules forced the reduced usage of CUG—and also CUA, probably as a result of the weaker wobble—in *Candida* genes (Fig. 3a). CUU and CUC codons do not display the same bias for infrequent usage

**Table 1 | *Candida* genome features**

| Species* | Genome size (Mb) | GC content (%) | No. of genes | Ave. gene size (bp) | Intergenic ave. (bp) | Ploidy | Pathogen† |
|---|---|---|---|---|---|---|---|
| *C. albicans* WO-1 | 14.4 | 33.5 | 6,159 | 1,444 | 921 | diploid | ++ |
| *C. albicans* SC5314 | 14.3 | 33.5 | 6,107 | 1,468 | 858 | diploid | ++ |
| *C. tropicalis* | 14.5 | 33.1 | 6,258 | 1,454 | 902 | diploid | ++ |
| *C. parapsilosis* | 13.1 | 38.7 | 5,733 | 1,533 | 752 | diploid | ++ |
| *L. elongisporus* | 15.4 | 37.0 | 5,802 | 1,530 | 1,174 | diploid | − |
| *C. guilliermondii* | 10.6 | 43.8 | 5,920 | 1,402 | 426 | haploid | + |
| *C. lusitaniae* | 12.1 | 44.5 | 5,941 | 1,382 | 770 | haploid | + |
| *D. hansenii* | 12.2 | 36.3 | 6,318 | 1,382 | 550 | haploid | − |

bp, base pair.
\* *C. albicans* SC5314 assembly 21 and gene set dated 28 January 2008 downloaded from the *Candida* Genome Database; *D. hansenii* assembly from GenBank[9]. The remaining assemblies are reported as part of this work, and are available in GenBank and at the Broad Institute *Candida* Database website.
† Relative level of pathogen strength: ++, strong pathogen; +, moderate pathogen; −, rare pathogen.
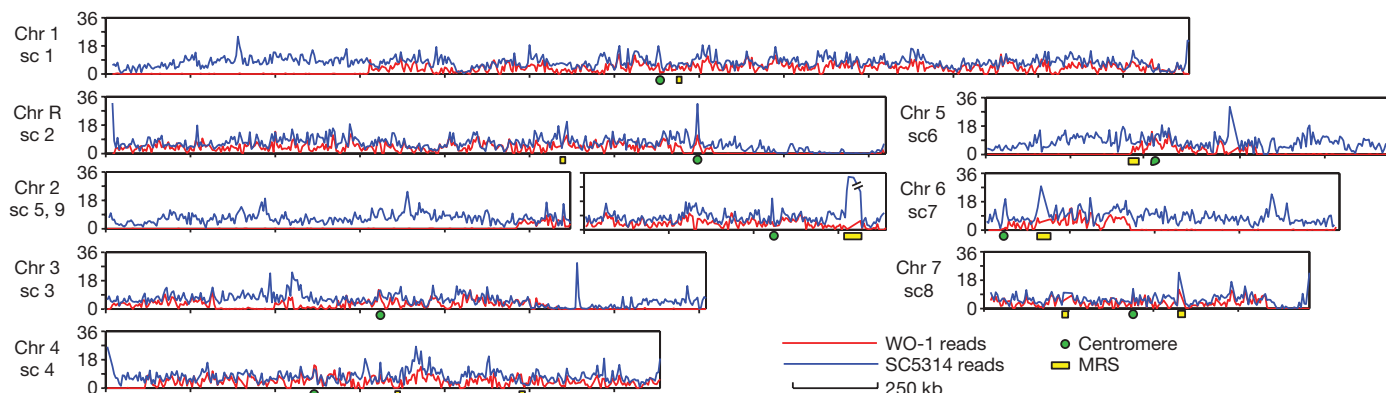
**Figure 2 | C. albicans WO-1 is highly homozygous.** Red lines show SNPs per kilobase, normalized by coverage, within WO-1, and blue lines show SNPs per kilobase between WO-1 and SC5314. Although both copies of chromosome 5 have rearranged at the MRS (yellow box) in WO-1, we show this as a single chromosome to allow a haploid reference for polymorphism. Relative to SC5314, chromosomes 1, 4 and 6 have the opposite orientation (Supplementary Fig. 6). Chr, chromosome; sc, supercontig; kb, kilobase.

(Fig. 3b). An additional pressure influencing codon usage may be the GC content, as usage of leucine codons in *Candida* species is correlated with the percentage GC composition (Supplementary Table 11).

We also examined the evolutionary fate of ancestral CUG codons and the origin of new CUG codons (Supplementary Table 12). CUG codons in *C. albicans* almost never (1%) align opposite CUG codons in *S. cerevisiae*. Instead, CUG serine codons in *C. albicans* align primarily to *Saccharomyces* codons for serine (20%) and other hydrophilic residues (49%). CUG leucine codons in *S. cerevisiae* align primarily to leucine codons in *Candida* (50%) and to other hydrophobic-residue codons (30%). This suggests a complete functional replacement of CUG codons in *Candida*.

### Gene family evolution

To identify gene families likely to be associated with *Candida* pathogenicity and virulence, we used a phylogenomic approach across



**Figure 3 | Evolutionary effects of CUG coding. a**, Average percentage of genes with zero, one or two CUG and CUA codons in *Candida* (grey bars) and *Saccharomyces* (black bars). Error bars, s.d. All differences are significant with $P \le 0.0004$ (*t*-test, $N = 8$ *Candida* spp. and 6 *Saccharomyces* spp.) except for those in genes with two CUA codons ($P = 0.02$). 'Candida' and 'Saccharomyces' here refer to the CTG and WGD clades in Fig. 1, but including *Pichia stipitis* and excluding *C. dubliniensis*. **b**, CUN codon usage for all codon counts. **c**, Decoding rules for CUN codons in *Saccharomyces* and *Candida*.

seven *Candida* and nine *Saccharomyces* genomes (Supplementary Information, section S10). Among 9,209 gene families, we identified 21 that are significantly enriched in the more common pathogens (Table 2). These include families encoding lipases, oligopeptide transporters and adhesins, which are all known to be associated with pathogenicity[8], as well as poorly characterized families not previously associated with pathogenesis.

Three cell wall families are enriched in the pathogens: those encoding Hyr/Iff proteins (ref. 16), Als adhesins[17] and Pga30-like proteins (Table 2 and Supplementary text S11). The Als family (family 17 in Supplementary Tables 22 and 23) in *C. albicans* is associated with virulence, and in particular with adhesion to host surfaces[18], invasion of host cells[19] and iron acquisition[20]. All these families are absent from the *Saccharomyces* clade species and are particularly enriched in the more pathogenic species (Supplementary Table 18). All three families are highly enriched for gene duplications (Supplementary Table 19), including tandem clusters of two to six genes, and show high mutation rates (fastest 5% of families) (Supplementary Information, section S10d). This variable repertoire of cell wall proteins is likely to be of profound importance to the niche adaptations and relative virulence of these organisms.

Als[17] and Hyr/Iff genes frequently contain intragenic tandem repeats, which modulate adhesion and biofilm formation in *S. cerevisiae*[21] (Fig. 4 and Supplementary Figs 19 and 20). The sequence of intragenic tandem repeats is conserved at the protein level across species (Supplementary Figs 19 and 20). Two proteins contain both an Als domain and repeats characteristic of the Hyr/Iff family.

*Candida* clade pathogens show expansions of extracellular enzyme and transmembrane transporter families (Table 2 and Supplementary Table 22). These families are either not found in *Saccharomyces* (including amino-acid permeases, lipases and superoxide dismutases) or are present in *S. cerevisiae* but significantly expanded in pathogens (including phospholipase B, ferric reductases, sphingomyelin phosphodiesterases and GPI-anchored yapsin proteases, which have been linked to virulence in *C. glabrata*[22]). Several groups of cell-surface transporters are also enriched (including oligopeptide transporters, amino-acid permeases and the major facilitator superfamily). Overall, these family expansions illustrate the importance of extracellular activities in virulence and pathogenicity. Genes involved in stress response are also variable between species (Supplementary Information, section S12).

*C. albicans* also showed species-specific expansion of some families, including two associated with filamentous growth, a leucine-rich repeat family and the Fgr6-1 family (Table 2). *Candida albicans* forms hyphae whereas *C. tropicalis* and *C. parapsilosis* produce only pseudohyphae, so these families may contribute to differences in hyphal growth.

**659**

**Table 2 | Gene families enriched in pathogenic *Candida* spp.**

| No. | Annotation | Pathogen genes | Non-pathogen genes | P val. | Dup. | Loss | Gene rate | C. alb. | C. tro. | C. par. | L. elo. | C. gui. | C. lus. | D. han. | C. gla. | Yeast (ave.) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | GPI family 18 (Hyr/Iff-like) | 56 | 10 | $1.4 \times 10^{-16}$ | 52 | 11 | 16.2 | 11 | 18 | 17 | 9 | 3 | 7 | 1 | 0 | 0.0 |
| 2 | Leucine-rich repeat (Ifa/Fgr38-like) | 34 | 0 | $4.2 \times 10^{-16}$ | 32 | 5 | 18.3 | 33 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 |
| 3 | Ferric reductase family | 45 | 10 | $1.9 \times 10^{-12}$ | 30 | 25 | 2.5 | 12 | 19 | 7 | 7 | 3 | 4 | 2 | 0 | 0.1 |
| 4 | Reductase family | 43 | 11 | $3.2 \times 10^{-11}$ | 31 | 30 | 2.3 | 7 | 12 | 9 | 6 | 13 | 2 | 4 | 0 | 0.1 |
| 5 | GPI family 17 (Als-like adhesins) | 31 | 5 | $4.4 \times 10^{-10}$ | 29 | 4 | 20.5 | 8 | 16 | 5 | 4 | 2 | 0 | 1 | 0 | 0.0 |
| 6 | GPI family 13 (Pga30-like) | 34 | 7 | $5.0 \times 10^{-10}$ | 25 | 5 | 14.8 | 12 | 14 | 6 | 6 | 1 | 1 | 1 | 0 | 0.0 |
| 7 | Unclassified | 20 | 0 | $9.0 \times 10^{-10}$ | 13 | 3 | 15.9 | 9 | 9 | 0 | 0 | 2 | 0 | 0 | 0 | 0.0 |
| 8 | Cell wall mannoprotein biosynthesis | 38 | 18 | $7.2 \times 10^{-7}$ | 19 | 34 | 2.1 | 8 | 7 | 8 | 8 | 11 | 4 | 9 | 0 | 0.1 |
| 9 | Major facilitator transporters | 25 | 7 | $9.2 \times 10^{-7}$ | 14 | 17 | 2.0 | 3 | 3 | 7 | 3 | 10 | 2 | 4 | 0 | 0.0 |
| 10 | Oligopeptide transporters | 31 | 13 | $2.2 \times 10^{-6}$ | 23 | 11 | 6.7 | 6 | 9 | 9 | 4 | 4 | 3 | 1 | 0 | 0.9 |
| 11 | Unclassified | 25 | 9 | $6.3 \times 10^{-6}$ | 15 | 6 | 11.1 | 7 | 9 | 3 | 5 | 3 | 1 | 4 | 2 | 0.2 |
| 12 | Amino-acid permeases | 27 | 11 | $7.7 \times 10^{-6}$ | 11 | 18 | 1.7 | 6 | 6 | 6 | 4 | 6 | 3 | 6 | 0 | 0.1 |
| 13 | Sphingomyelin phosphodiesterases | 18 | 5 | $3.2 \times 10^{-5}$ | 11 | 9 | 7.4 | 4 | 5 | 4 | 2 | 3 | 2 | 1 | 0 | 0.2 |
| 14 | Fgr6 family (filamentous growth) | 12 | 1 | $3.3 \times 10^{-5}$ | 7 | 1 | 14.5 | 8 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0.0 |
| 15 | Secreted lipases | 20 | 7 | $4.6 \times 10^{-5}$ | 17 | 8 | 9.6 | 10 | 5 | 4 | 4 | 1 | 0 | 3 | 0 | 0.0 |
| 16 | Cytochrome P450 family | 34 | 21 | $5.5 \times 10^{-5}$ | 23 | 22 | 6.0 | 6 | 8 | 10 | 7 | 5 | 4 | 6 | 1 | 1.0 |
| 17 | Amino-acid permeases | 16 | 4 | $5.6 \times 10^{-5}$ | 14 | 10 | 1.5 | 2 | 3 | 6 | 3 | 2 | 3 | 1 | 0 | 0.0 |
| 18 | Zinc-finger transcription factors | 31 | 18 | $6.2 \times 10^{-5}$ | 17 | 14 | 12.3 | 5 | 8 | 7 | 7 | 7 | 4 | 11 | 0 | 0.0 |
| 19 | Unclassified | 13 | 2 | $6.3 \times 10^{-5}$ | 8 | 0 | 8.1 | 3 | 1 | 6 | 1 | 2 | 1 | 1 | 0 | 0.0 |
| 20 | Predicted transmembrane family | 17 | 5 | $7.2 \times 10^{-5}$ | 9 | 2 | 7.5 | 4 | 4 | 5 | 3 | 3 | 1 | 2 | 0 | 0.0 |
| 21 | Unclassified secreted family | 20 | 8 | $1.1 \times 10^{-4}$ | 7 | 6 | 9.3 | 4 | 4 | 6 | 4 | 4 | 2 | 4 | 0 | 0.0 |

Pathogen genes: total genes in family for *C. albicans*, *C. tropicalis*, *C. parapsilosis*, *C. guilliermondii*, *C. lusitaniae* and *C. glabrata*. Non-pathogen genes: total genes in family for *L. elongisporus*, *D. hansenii* and all *Saccharomyces* clade species (Fig. 1) except *C. glabrata*. P val., P value of the hypergeometric test; all families shown above have a false discovery rate of less than 0.05 (Supplementary Information, section S10c). Dup., duplications; Loss, losses (Supplementary Information, section S10b). Gene rate: average mutation rate for each family (Supplementary Information, section S10d); the average gene rate across all families is 5.8. Yeast (ave.): average count for all *Saccharomyces* species. GPI, glycosyl phosphatidylinositol.

We identified 64 families showing positive selection in the highly pathogenic *Candida* species (Supplementary Table 32). These are highly enriched for cell wall, hyphal, pseudohyphal, filamentous growth and biofilm functions (Supplementary Information, section S13). Six of the families have been previously associated with pathogenesis, including that of *ERG3*, a C-5 sterol desaturase essential for ergosterol biosynthesis, for which mutations can cause drug resistance[23].

## Structure of the *MTL* locus

Pathogenic fungi may have limited their sexual cycles to maximize their virulence[24], and the sequenced *Candida* species show tremendous diversity in their apparent abilities to mate. Among the four



**Figure 4 | Conserved domains of Als and Hyr/Iff cell wall families.** The amino-terminal Als domain (red) and Hyr/Iff domain (green) are shown as ovals. Intragenic tandem repeats (ITRs; see Supplementary Information, section S11) are shown as rectangles, coloured to represent similar amino-acid sequences.

diploids, *C. albicans* has a parasexual cycle (mating of diploid cells followed by mitosis and chromosome loss instead of meiosis[25]), *L. elongisporus* has been described as sexual and homothallic (self-mating)[26], whereas *C. tropicalis* and *C. parapsilosis* have never been observed to mate. Among the three haploids, *C. guilliermondii* and *C. lusitaniae* are heterothallic (cross-mating only) and have a complete sexual cycle, whereas *D. hansenii* is haploid and homothallic[27] (Supplementary Information, section 14c).

To understand the genomic basis for this diversity, we studied the *Candida MTL* locus, which determines mating type, similar to the *MAT* locus in *S. cerevisiae*. In both *C. albicans* and *S. cerevisiae*, the mating locus has two idiomorphs, **a** and α, encoding the regulators **a**1, α1 and α2, respectively. *Candida albicans MTL***a** also encodes **a**2, and both idiomorphs in this species contain alleles of three additional genes without known roles in mating: *PAP*, *OBP* and *PIK*[28]. The *MTL*α and *MTL***a** genes, alone or in combination, specify one of the three possible cell-type programs (**a** haploid, α haploid, **a**/α diploid). In *C. albicans*, the alpha-domain protein α1 activates α-specific mating genes, the high-mobility group (HMG) factor **a**2 activates **a**-specific mating genes, and the **a**1/α2 homeodomain heterodimer represses mating genes in **a**/α cells[29,30].

Despite extended conservation of the genomic context flanking the mating-type locus, there is great variability in *MTL* gene content (Fig. 5). *MTL***a**1 has become a pseudogene in *C. parapsilosis*[31], and is probably a recent loss because target genes retain predicted **a**1/α2 binding sites (Supplementary Information, section S14). *MTL*α2 is missing in both *C. guilliermondii* and *C. lusitaniae* (J. L. Reedy, A. Floyd and J. Heitman, submitted). A fused mating-type locus containing both **a** and α genes is found in *D. hansenii* and *Pichia stipitis*[32,33].

Most surprisingly, all four mating-type genes are missing in *L. elongisporus*. It contains a site syntenic to *MTL***a** in other species, but this contains only 508 base pairs of apparently non-coding DNA,
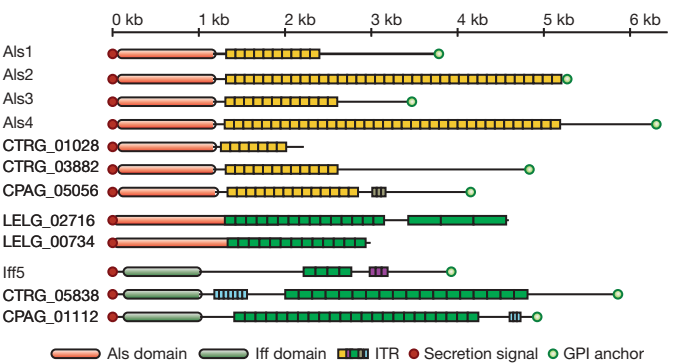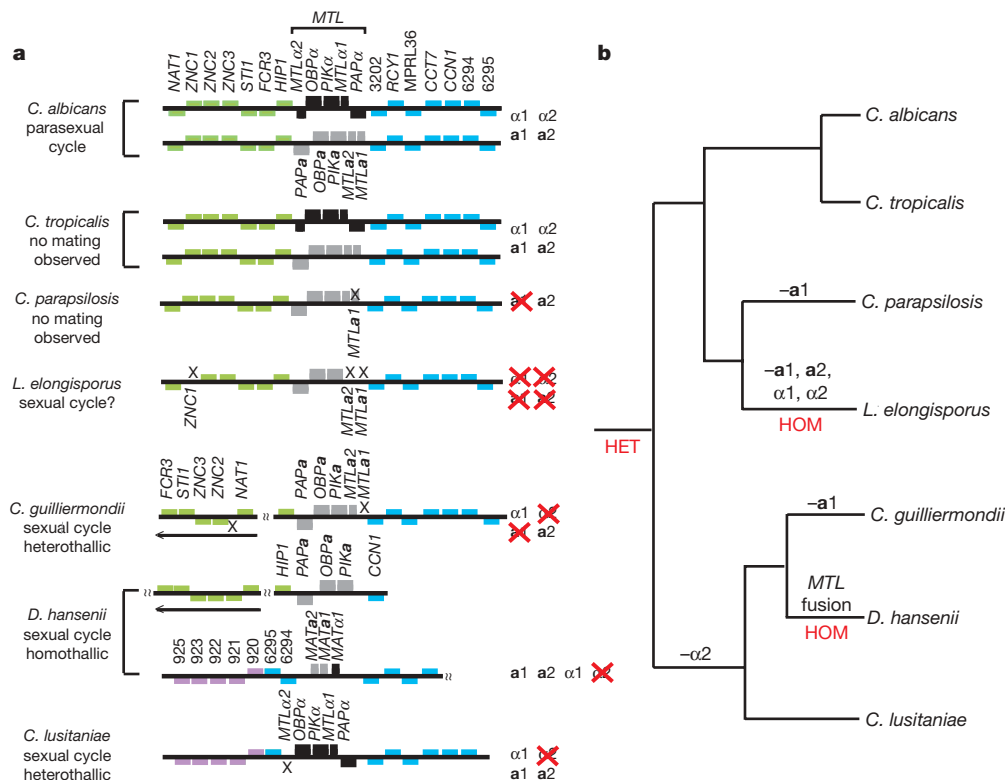
**Figure 5 | Organization of *MTL* loci in the *Candida* clade. a**, *MTL**a***-specific genes are shown in grey, *MTL*α-specific genes are shown in black and other orthologues are shown in colour. Two idiomorphs from *C. albicans* and *C. tropicalis* are shown. Arrows indicate inversions relative to *C. albicans*. Crosses show gene losses and the *C. parapsilosis MTL**a**1* pseudogene. There is no *MTL* locus in *L. elongisporus*; gene order around the *PAP**a***, *OBP**a*** and *PIK**a*** genes is shown. For *C. guilliermondii* and *C. lusitaniae*, the genome project sequenced one idiomorph; the second was obtained by J. Reedy. In *D. hansenii*, *PAP*, *OBP* and *PIK* are separate from the fused *MTL* locus[32]. **b**, Placement of gene losses on the phylogenetic tree. HOM, homothallic; HET, heterothallic.

a length insufficient to encode **a**1 or **a**2 even if they were extensively divergent in sequence. We confirmed this finding in seven other *L. elongisporus* isolates (not shown). The sexual state of *L. elongisporus* has been assumed to be homothallic because asci are generated from identical cells[5,26], but the absence of an **a**-factor pheromone, receptor and transporter (Supplementary Table 34) as well as *MTL*, suggests that it may not have a sexual cycle. Alternatively, mating may require only one pheromone and receptor (α), and *L. elongisporus* may be the first identified ascomycete that can mate independently of *MTL* or *MAT*.

Our discovery that *MTL*α2 and *MTL**a**1* are frequently absent challenges our understanding of how the mating locus operates. The model derived for mating regulation in *C. albicans*, including the role of **a**1/α2 in white–opaque switching[3], must differ substantially in the other sexual species. This is particularly interesting because the major regulators of the white–opaque switch in *C. albicans* (*WOR1*, *WOR2*, *CZF1* and *EFG1*[34]) are generally conserved in other species, but Wor1 control over white–opaque genes appears to be a recent innovation in *C. albicans* and *Candida dubliniensis*[35]. Our data also suggest that the loss of α2 occurred early in the haploid sexual lineage.

**Mating and meiosis**

To gain further insight into their diversity in sexual behaviour, we examined whether 227 genes required for meiosis in *S. cerevisiae* and other fungi have orthologues in the *Candida* species (Supplementary Information, section S14). A previous report[36] that some components of meiosis, such as the major regulator *IME1*, are missing from *C. albicans* led us to propose that their loss could be correlated with lack of meiosis. Surprisingly, however, we find that these genes are missing in all *Candida* species, suggesting that sexual *Candida* species undergo meiosis without them. Conversely, even seemingly non-mating species showed highly conserved pheromone response

pathways, suggesting that pheromone signalling plays an alternative role such as regulation of biofilm formation[37]. These findings suggest considerable plasticity and innovation of meiotic pathways in *Candida*.

Moreover, we find that sexual *Candida* species have undergone a recent dramatic change in the pathways involved in meiotic recombination, with loss of the Dmc1-dependent pathway in the heterothallic species *C. lusitaniae* and *C. guilliermondii* (Supplementary Information, section 14c). We also found that mechanisms of chromosome pairing and crossover formation have changed recently in these two species, because they (and to a lesser extent *D. hansenii*) have lost several components of the synaptonemal and synapsis initiation complexes (Supplementary Table 35). They have also lost components of the major crossover-formation pathway in *S. cerevisiae* (*MSH4*, *MSH5*), but have retained a minor pathway (*MUS81*, *MMS4*)[38,39]. Overall, if *Candida* species undergo meiosis it is with reduced machinery, or different machinery, suggesting that unrecognized meiotic cycles may exist in many species, and that the model of meiosis developed in *S. cerevisiae* varies significantly, even among yeasts.

The genome sequences reported here provide a resource that will allow current knowledge of *C. albicans* biology, the product of decades of research, to be applied with maximum effect to the other pathogenic species in the *Candida* clade. They also allow many of the unusual features of *C. albicans*—such as cell wall gene family amplifications, and its apparent ability to undergo mating and a parasexual cycle without meiosis—to be understood in an evolutionary context that shows that the genes involved in virulence and mating have highly dynamic rates of turnover and loss.

**METHODS SUMMARY**

The methods for this paper are described in Supplementary Information. Here we outline the resources generated by this project.

**661**

Assemblies, gene sets, and single nucleotide polymorphisms are available in GenBank and at the Broad Institute *Candida* Database website (http://www.broad.mit.edu/annotation/genome/candida_group/MultiHome.html). The Broad Institute website provides search, visualization, BLAST and download of assemblies and gene sets. Gene families can be accessed by searching either for individual genes or with family identifiers (CF#####). The *C. parapsilosis* assembly is also available at the Wellcome Trust Sanger Institute website (http://www.sanger.ac.uk/sequencing/Candida/parapsilosis/). The revised annotation of *C. albicans* (SC5314) is available at the *Candida* Genome Database (www.candidagenome.org).

1. Pfaller, M. A. & Diekema, D. J. Epidemiology of invasive candidiasis: a persistent public health problem. *Clin. Microbiol. Rev.* **20**, 133–163 (2007).
2. Santos, M. A. & Tuite, M. F. The CUG codon is decoded *in vivo* as serine and not leucine in *Candida albicans. Nucleic Acids Res.* **23**, 1481–1486 (1995).
3. Lockhart, S. R. *et al.* In *Candida albicans*, white-opaque switchers are homozygous for mating type. *Genetics* **162**, 737–745 (2002).
4. Slutsky, B., Buffo, J. & Soll, D. R. High-frequency switching of colony morphology in *Candida albicans. Science* **230**, 666–669 (1985).
5. Lockhart, S. R., Messer, S. A., Pfaller, M. A. & Diekema, D. J. *Lodderomyces elongisporus* masquerading as *Candida parapsilosis* as a cause of bloodstream infections. *J. Clin. Microbiol.* **46**, 374–376 (2008).
6. Braun, B. R. *et al.* A human-curated annotation of the *Candida albicans* genome. *PLoS Genet.* **1**, e1 (2005).
7. Jones, T. *et al.* The diploid genome sequence of *Candida albicans. Proc. Natl Acad. Sci. USA* **101**, 7329–7334 (2004).
8. van het Hoog, M, *et al.* Assembly of the *Candida albicans* genome into sixteen supercontigs aligned on the eight chromosomes. *Genome Biol.* **8**, R52 (2007).
9. Dujon, B. *et al.* Genome evolution in yeasts. *Nature* **430**, 35–44 (2004).
10. Zhang, J., Hollis, R. J. & Pfaller, M. A. Variations in DNA subtype and antifungal susceptibility among clinical isolates of *Candida tropicalis. Diagn. Microbiol. Infect. Dis.* **27**, 63–67 (1997).
11. Tavanti, A. *et al.* Population structure and properties of *Candida albicans*, as determined by multilocus sequence typing. *J. Clin. Microbiol.* **43**, 5601–5613 (2005).
12. Chu, W. S., Magee, B. B. & Magee, P. T. Construction of an SfiI macrorestriction map of the *Candida albicans* genome. *J. Bacteriol.* **175**, 6637–6651 (1993).
13. Forche, A., Magee, P. T., Magee, B. B. & May, G. Genome-wide single-nucleotide polymorphism map for *Candida albicans. Eukaryot. Cell* **3**, 705–714 (2004).
14. Legrand, M. *et al.* Haplotype mapping of a diploid non-meiotic organism using existing and induced aneuploidies. *PLoS Genet.* **4**, e1 (2008).
15. Massey, S. E. *et al.* Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in *Candida* spp. *Genome Res.* **13**, 544–557 (2003).
16. Bates, S. *et al. Candida albicans* Iff11, a secreted protein required for cell wall structure and virulence. *Infect. Immun.* **75**, 2922–2928 (2007).
17. Hoyer, L. L., Green, C. B., Oh, S. H. & Zhao, X. Discovering the secrets of the *Candida albicans* agglutinin-like sequence (ALS) gene family–a sticky pursuit. *Med. Mycol.* **46**, 1–15 (2008).
18. Yeater, K. M. *et al.* Temporal analysis of *Candida albicans* gene expression during biofilm development. *Microbiology* **153**, 2373–2385 (2007).
19. Phan, Q. T. *et al.* Als3 is a *Candida albicans* invasin that binds to cadherins and induces endocytosis by host cells. *PLoS Biol.* **5**, e64 (2007).
20. Almeida, R. S. *et al.* The hyphal-associated adhesin and invasin Als3 of *Candida albicans* mediates iron acquisition from host ferritin. *PLoS Pathog.* **4**, e1000217 (2008).
21. Verstrepen, K. J., Jansen, A., Lewitter, F. & Fink, G. R. Intragenic tandem repeats generate functional variability. *Nature Genet.* **37**, 986–990 (2005).
22. Kaur, R., Ma, B. & Cormack, B. P. A family of glycosylphosphatidylinositol-linked aspartyl proteases is required for virulence of *Candida glabrata. Proc. Natl Acad. Sci. USA* **104**, 7628–7633 (2007).
23. Chau, A. S. *et al.* Inactivation of sterol $\Delta^{5,6}$-desaturase attenuates virulence in *Candida albicans. Antimicrob. Agents Chemother.* **49**, 3646–3651 (2005).
24. Nielsen, K. & Heitman, J. Sex and virulence of human pathogenic fungi. *Adv. Genet.* **57**, 143–173 (2007).
25. Noble, S. M. & Johnson, A. D. Genetics of *Candida albicans*, a diploid human fungal pathogen. *Annu. Rev. Genet.* **41**, 193–211 (2007).
26. van der Walt, J. P. *Lodderomyces*, a new genus of the Saccharomycetaceae. *Antonie Van Leeuwenhoek* **32**, 1–5 (1966).
27. van der Walt, J. P., Taylor, M. B. & Liebenberg, N. V. D. W. Ploidy, ascus formation and recombination in *Torulaspora* (*Debaryomyces*) *hansenii. Antonie Van Leeuwenhoek* **43**, 205–218 (1977).
28. Hull, C. M., Raisner, R. M. & Johnson, A. D. Evidence for mating of the ''asexual'' yeast *Candida albicans* in a mammalian host. *Science* **289**, 307–310 (2000).
29. Tsong, A. E., Miller, M. G., Raisner, R. M. & Johnson, A. D. Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* **115**, 389–399 (2003).
30. Tsong, A. E., Tuch, B. B., Li, H. & Johnson, A. D. Evolution of alternative transcriptional circuits with identical logic. *Nature* **443**, 415–420 (2006).
31. Logue, M. E., Wong, S., Wolfe, K. H. & Butler, G. A genome sequence survey shows that the pathogenic yeast *Candida parapsilosis* has a defective MTLa1 allele at its mating type locus. *Eukaryot. Cell* **4**, 1009–1017 (2005).
32. Fabre, E. *et al.* Comparative genomics in hemiascomycete yeasts: evolution of sex, silencing, and subtelomeres. *Mol. Biol. Evol.* **22**, 856–873 (2005).
33. Jeffries, T. W. *et al.* Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis. Nature Biotechnol.* **25**, 319–326 (2007).
34. Zordan, R. E., Miller, M. G., Galgoczy, D. J., Tuch, B. B. & Johnson, A. D. Interlocking transcriptional feedback loops control white-opaque switching in *Candida albicans. PLoS Biol.* **5**, e256 (2007).
35. Tuch, B. B., Galgoczy, D. J., Hernday, A. D., Li, H. & Johnson, A. D. The evolution of combinatorial gene regulation in fungi. *PLoS Biol.* **6**, e38 (2008).
36. Tzung, K. W. *et al.* Genomic evidence for a complete sexual cycle in *Candida albicans. Proc. Natl Acad. Sci. USA* **98**, 3249–3253 (2001).
37. Daniels, K. J., Srikantha, T., Lockhart, S. R., Pujol, C. & Soll, D. R. Opaque cells signal white cells to form biofilms in *Candida albicans. EMBO J.* **25**, 2240–2252 (2006).
38. Argueso, J. L., Wanat, J., Gemici, Z. & Alani, E. Competing crossover pathways act during meiosis in *Saccharomyces cerevisiae. Genetics* **168**, 1805–1816 (2004).
39. de los Santos, T. *et al.* The Mus81/Mms4 endonuclease acts independently of double-Holliday junction resolution to promote a distinct subset of crossovers during meiosis in budding yeast. *Genetics* **164**, 81–94 (2003).
40. Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S. & Wolfe, K. H. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* **440**, 341–345 (2006).

# ARTICLES

# Driving fast-spiking cells induces gamma rhythm and controls sensory responses

Jessica A. Cardin[1,2]*, Marie Carlén[3,4]*, Konstantinos Meletis[3,4], Ulf Knoblich[1], Feng Zhang[5], Karl Deisseroth[5], Li-Huei Tsai[3,4,6] & Christopher I. Moore[1]

Cortical gamma oscillations (20−80 Hz) predict increases in focused attention, and failure in gamma regulation is a hallmark of neurological and psychiatric disease. Current theory predicts that gamma oscillations are generated by synchronous activity of fast-spiking inhibitory interneurons, with the resulting rhythmic inhibition producing neural ensemble synchrony by generating a narrow window for effective excitation. We causally tested these hypotheses in barrel cortex *in vivo* by targeting optogenetic manipulation selectively to fast-spiking interneurons. Here we show that light-driven activation of fast-spiking interneurons at varied frequencies (8−200 Hz) selectively amplifies gamma oscillations. In contrast, pyramidal neuron activation amplifies only lower frequency oscillations, a cell-type-specific double dissociation. We found that the timing of a sensory input relative to a gamma cycle determined the amplitude and precision of evoked responses. Our data directly support the fast-spiking-gamma hypothesis and provide the first causal evidence that distinct network activity states can be induced *in vivo* by cell-type-specific activation.

Brain states characterized by rhythmic electrophysiological activity have been studied intensively for more than 80 years[1,2]. Because these brain rhythms are believed to be essential to information processing, many theories have been proposed to explain their origin, with several emphasizing the activity of neural subtypes. One of the strongest cases made so far for the importance of a specific cell type in rhythm induction is the suggested role of fast-spiking (FS) interneurons in gamma oscillations[3–6]. Networks of FS cells connected by gap junctions[7,8] provide large, synchronous inhibitory postsynaptic potentials (IPSPs) to local excitatory neurons[9,10]. Computational modelling indicates that this synchronous activity is sufficient to induce 20−80-Hz oscillations that are stabilized and regulated by fast excitatory feedback from pyramidal neurons[11,12]. Cortical recordings *in vivo* show sensory-evoked gamma oscillations in the local field potential (LFP) and phase-locked firing of excitatory pyramidal cells, indicating entrainment of excitatory neurons to rhythmic inhibitory activity[9,13–15]. Despite considerable study of cortical oscillations and the importance of understanding their origins, induction of a given network state by stimulation of specific neural cell types *in vivo* has not previously been possible.

## Cell-type-specific expression of channelrhodopsin-2

To test directly the hypothesis that FS interneuron activity in an *in vivo* cortical circuit is sufficient to induce gamma oscillations, we used the light-sensitive bacteriorhodopsin *Chlamydomonas reinhardtii* channelrhodopsin-2 (ChR2), a cation channel activated by ∼470 nm blue light[16,17]. We targeted expression of ChR2 specifically to parvalbumin-positive fast-spiking (FS-PV⁺) interneurons by injecting the adeno-associated viral vector double-floxed inverted open reading frame *ChR2−mCherry* (AAV DIO *ChR2−mCherry*), with Cre-dependent expression of ChR2, into PV−Cre knock-in mice (Fig. 1a, Supplementary Figs 1 and 2 and Methods)[18,19]. Six days after virus injection into barrel cortex of adult PV−Cre mice, ChR2−mCherry expression covered an anterioposterior distance of up to 1,740 μm

(1,695 ± 57.4 μm, mean ± s.d., $n = 3$), resulting in robust labelling of PV⁺ interneurons across cortical layers (Fig. 1b). The labelling efficiency of AAV DIO *ChR2−mCherry* varied over distance from the injection site; close to the centre of the injection, >97% of the PV⁺ interneurons expressed ChR2−mCherry. Immunohistochemistry confirmed that 96.7 ± 1.0% (mean ± s.d., $n = 4,234$ ChR2−mCherry⁺ neurons, 4 animals) of the ChR2−mCherry⁺ neurons expressed PV (Fig. 1d, e and Supplementary Fig. 2), and almost all expressed the inhibitory neurotransmitter GABA (γ-aminobutyric acid; Supplementary Fig. 3)[20–23]. Expression of ChR2−mCherry was not induced after injection of AAV DIO *ChR2−mCherry* into wild-type mice (data not shown) or *in vitro* in the absence of Cre (see Supplementary Methods; data not shown).

In experiments targeting excitatory neurons, AAV DIO *ChR2−mCherry* was injected into the barrel cortex of adult CW2 (ref. 24) mice, which express Cre from the αCamKII (also known as Camk2a) promoter ('αCamKII−Cre mice'), inducing recombination in excitatory neurons in cortex[24]. Robust ChR2−mCherry expression was observed in excitatory neurons in a laminar profile corresponding to the Cre expression pattern (Fig. 1c and Supplementary Fig. 4)[24]. At least 50% of the αCamKII⁺ neurons in layer 2/3 expressed ChR2−mCherry (913 of 1,638 cells in a total area of $8.4 \times 10^6 \mu m^3$) close to the injection site, covering an anterioposterior distance of 1,560 ± 154.9 μm (mean ± s.d., $n = 3$). Immunohistochemical analysis revealed that 100 ± 0% (mean ± s.d., $n = 4,024$ ChR2−mCherry⁺ neurons, 4 animals) of the ChR2−mCherry-expressing neurons were immuno-negative for PV (Fig. 1f, g and Supplementary Fig. 2), and 100 ± 0% expressed the neuronal marker NeuN (data not shown).

## FS activation suppresses local sensory responses

We recorded light-activated FS and regular spiking (RS) single units in layers 2/3 and 4 of barrel cortex in PV−Cre ($n = 64$ FS cells in 15 animals) and αCamKII−Cre ($n = 56$ RS cells in 7 animals) mice. We

[1]McGovern Institute for Brain Research and Department of Brain and Cognitive Sciences, MIT, Cambridge, Massachusetts 02139, USA. [2]Department of Neuroscience, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [3]Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, MIT, Cambridge, Massachusetts 02139, USA. [4]Stanley Center for Psychiatric Research, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [5]Department of Bioengineering, Stanford University, Stanford, California 94305, USA. [6]Howard Hughes Medical Institute, Cambridge, Massachusetts 02139, USA.
*These authors contributed equally to this work.

663

**a** AAV DIO ChR2−mCherry



**b**



**c**
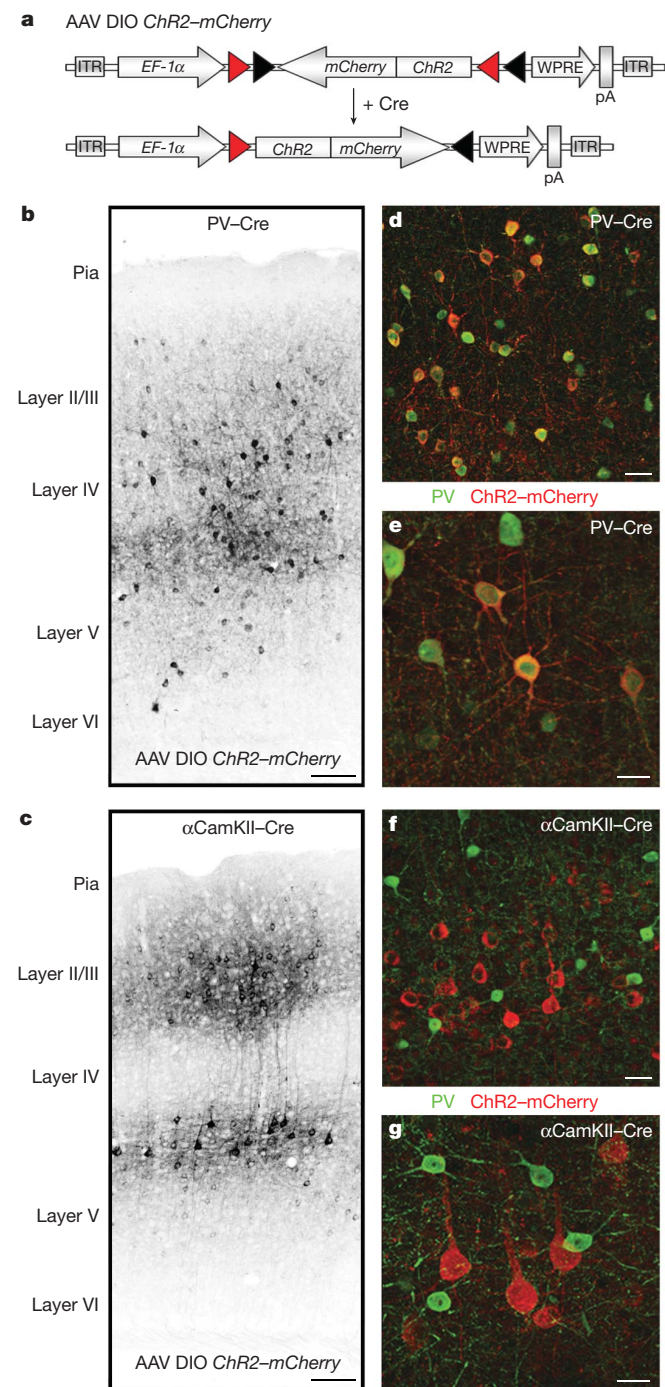


**Figure 1 | AAV DIO *ChR2−mCherry* gives Cre-dependent and cell-type-specific expression of light-activated channels *in vivo*. a,** AAV DIO *ChR2−mCherry* with Cre-dependent expression of ChR2 produced cell-type-specific targeting of light-activated channels. In the presence of Cre, *ChR2−mCherry* is inverted into the sense direction and expressed from the *EF-1α* (*EEF1A1*) promoter. ITR, inverted terminal repeat; pA, poly(A); WPRE, woodchuck hepatatis B virus post-transcriptional element. **b,** ChR2−mCherry was robustly expressed in PV$^+$ interneurons in barrel cortex of adult PV−Cre mice. **c,** A corresponding injection in αCamKII−Cre mice resulted in exclusive labelling of excitatory neurons. **d, e,** ChR2−mCherry expression in PV−Cre mice was confined to cells expressing PV. **e,** PV$^+$ cells with ChR2−mCherry expression and typical FS interneuron morphology. **f, g,** ChR2−mCherry expression in αCamKII−Cre mice is confined to neurons immuno-negative for PV. **g,** ChR2−mCherry-expressing cells with typical pyramidal neuron morphology. Scale bars: **b, c,** 100 μm; **d–g,** 25 μm.

did not observe light activation of layer 5 FS cells ($n = 12$ sites in 7 animals). Barrel cortex, which processes information from the rodent vibrissae (whiskers), was targeted as a well-defined model of basic sensory cortical function. In agreement with the immunohistological results, the action potential shapes of the neurons activated by light pulses were differentiated into two discrete populations based on mouse type: PV−Cre/FS and αCamKII−Cre/RS ($P < 0.01$; Fig. 2a).

To confirm the activation of inhibitory interneurons and their postsynaptic impact on excitatory neurons, we performed *in vivo* intracellular recordings of RS cells in barrel cortex of PV−Cre mice ($n = 5$). We found that a 1-ms light pulse was sufficient to evoke large, fast IPSPs, confirming direct synaptic inhibition of RS cells by light-activated FS cells (Fig. 2b). The latencies of the presynaptic light-evoked FS spikes agreed well with the onset times of the post-synaptic IPSPs, with FS spikes preceding IPSP onset by 0.5 to 0.75 ms (Fig. 2c). Both the time to peak and the peak timing variability of the evoked IPSPs decreased with increasing light pulse power (Fig. 2c). Mean IPSP peak amplitude at membrane potentials of −55 to −60 mV was 2.7 ± 1.0 mV. The mean reversal potential of the evoked IPSPs (see Supplementary Methods) was –67.6 ± 1.9 mV, indicating a GABA$_A$-mediated Cl$^−$ conductance characteristic of FS synapses. Consistent with IPSP induction, activation of FS cells blocked vibrissa-evoked responses in neighbouring RS cells (Fig. 2d, e; $n = 6$ sites in 5 PV−Cre mice).

## FS activation generates gamma oscillations

A strong prediction of the FS-gamma hypothesis is that synchronously active FS cells are sufficient for gamma induction. This hypothesis
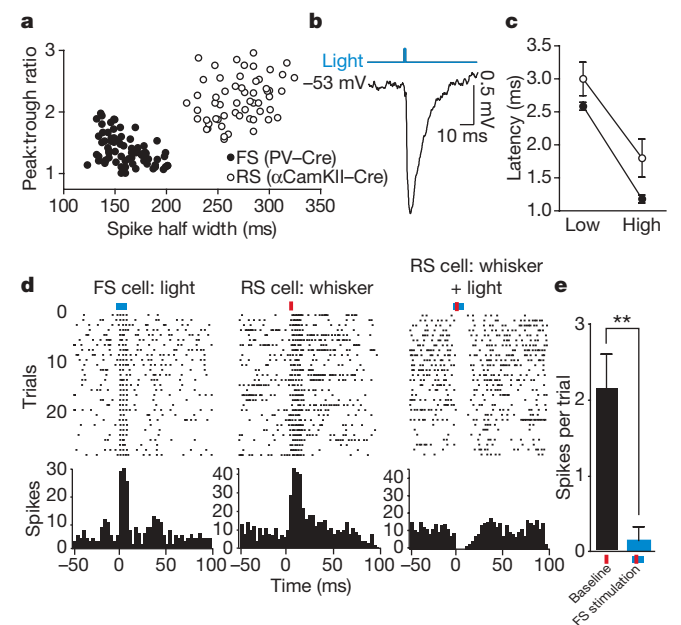


**Figure 2 | Light-evoked activity in FS-PV$^+$ inhibitory interneurons suppresses sensory processing in nearby excitatory neurons. a,** Light-activated RS and FS cells recorded in layers 2/3 and 4 of barrel cortex in PV−Cre and αCamKII−Cre mice, respectively, formed two discrete overall populations based on waveform properties. **b,** Intracellular *in vivo* recording of an RS cell in a PV−Cre animal. A 1-ms pulse of blue light at low power evoked an IPSP with a sharp onset. **c,** The latency to light-activated FS spikes (filled circles) agreed well with the onset latency of the resulting IPSPs (open circles). The IPSP time to peak decreased with increasing power (low power: 46 mW mm$^{-2}$; high power: 68 mW mm$^{-2}$). **d,** Sustained activation of FS inhibitory interneurons eliminated sensory responses in nearby RS neurons. A layer 2/3 FS cell was reliably activated by a 10-ms light pulse (blue line; left panel). An RS cell recorded on the same tetrode responded to vibrissa deflection (red bar; centre panel). Activation of inhibitory activity simultaneously with vibrissa deflection eliminated the RS sensory response (right panel). **e,** Mean RS vibrissa response decreased significantly in the presence of increased FS cell activity. **$P < 0.01$; error bars, mean ± s.e.m.

predicts that light pulses presented at a broad range of frequencies should reveal a selective peak in enhancement of the LFP, a measure of synchronous local network activity[25], when FS cells are driven in the gamma range.

To test this hypothesis, we drove cortical FS cell spiking in virus-transduced PV−Cre mice at a range of frequencies (8−200 Hz) with 1-ms light pulses. Light pulses in the gamma range (40 Hz) resulted in reliable action potential output at 25-ms intervals (Fig. 3a). Across the population, FS and RS cells were driven with equally high reliability by light pulses at low frequencies (Fig. 3b). At higher frequencies, spike probability on each light cycle remained high for FS cells but decreased for RS cells.

Driving FS cells at 40 Hz caused a specific increase in the 35−40 Hz frequency band in the LFP (Fig. 3c and Supplementary Figs 5 and 6). We found that activation of FS cells in the 20−80 Hz range resulted in significant amplification of LFP power at those frequencies ($n = 14$ sites in 6 animals; Fig. 3d). However, activation of FS cells at lower frequencies did not affect LFP power, despite robust evoked FS firing on every light cycle. In contrast, 8−24 Hz light activation of RS cells in αCamKII−Cre mice induced increased LFP power at these frequencies, but RS activation at higher frequencies did not affect LFP power ($n = 13$ sites in 5 mice; Fig. 3d and Supplementary Fig. 5). Light stimulation in the untransduced contralateral barrel cortex did not affect LFP power at any frequency ($n = 6$ PV−Cre and 5 αCamKII−Cre animals; Supplementary Fig. 6).

This double dissociation of cell-type-specific state induction (gamma by FS and lower frequencies by RS) directly supports the prediction that FS-PV[+] interneuron activation is sufficient and specific for induction of gamma oscillations. To highlight this distinction, we compared the effects of stimulating the two cell types at 8 and 40 Hz. Stimulation of FS cells at 8 Hz in the PV−Cre mice had no effect on LFP power at 8 Hz, but FS stimulation at 40 Hz caused a significant increase in 40-Hz LFP power (paired t-test; $P < 0.001$; Fig. 3e). In contrast, stimulation of RS cells at 8 Hz in the αCamKII−Cre mice caused a significant increase of LFP power at 8 Hz ($P < 0.001$), whereas RS stimulation at 40 Hz caused only a small, nonsignificant increase in 40-Hz LFP power (Fig. 3f).

## Gamma generation is a resonant circuit property

One possible explanation for these results is that increased FS firing recruits resonant gamma-range activity in the surrounding local network as a function of the synaptic and biophysical properties of the cortical circuit. Alternatively, the increase in gamma activity may result from the specific level of evoked FS spiking, and changing spiking probability would shift the frequency of the enhanced LFP band. To discriminate between these possibilities, we stimulated FS cells at varying levels of light intensity. We found that FS spike probability changed with light intensity such that the spike probability curve shifted laterally (Fig. 3g). Whereas drive affected the amplitude of enhancement, LFP power was selectively amplified within the gamma
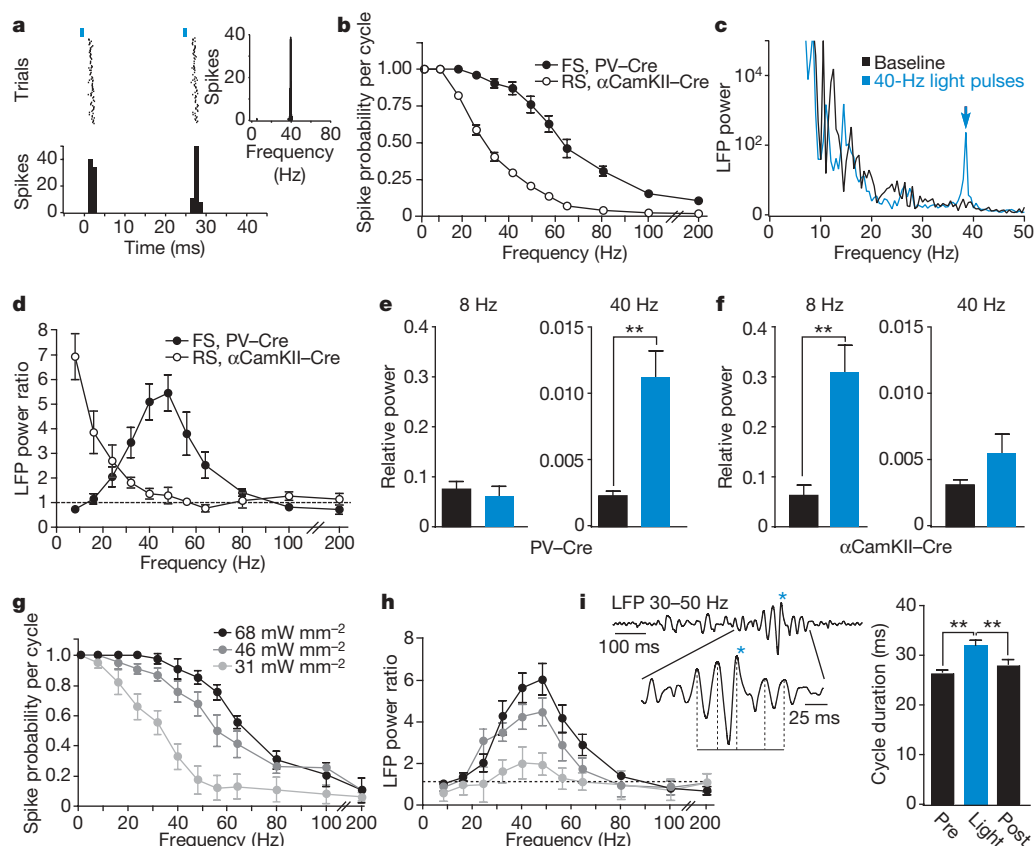


**Figure 3 | FS inhibitory interneurons generate gamma oscillations in the local cortical network. a**, In response to 40-Hz light pulses (blue bars), this FS cell fired reliably at 25-ms intervals, giving an instantaneous firing frequency of 40 Hz (inset). **b**, Average spike probability per light-pulse cycle in light-activated FS and RS cells in the PV−Cre and αCamKII−Cre mice, respectively (RS, $n = 17$, open circles; FS, $n = 22$, filled circles). **c**, Example of the increase in power at ~40 Hz in the LFP caused by activation of FS cells by light pulses at 40 Hz. **d**, Mean power ratio in each frequency band in response to light activation of FS (filled circles) and RS (open circles) cells at those frequencies. **e, f**, Comparison of the effect of activating FS and RS cells at 8 and 40 Hz on relative LFP power in those frequency bands. Black bars,

relative power in the baseline LFP; blue bars, relative power in the presence of light pulses. **g**, Average spike probability of FS cells per light pulse cycle in response to three levels of light intensity. **h**, Mean power ratios from LFP recordings at the light intensity levels shown in **g**. **i**, The trace shows spontaneously occurring gamma activity in the LFP. Brief activation of FS cells (blue asterisk) prolonged the duration of the ongoing gamma cycle and consequently shifted the phase of the following cycles. The duration of the cycle during which the light stimulus was given (Light) was significantly longer than the preceding (Pre) or the following (Post) cycle. **$P < 0.01$; error bars, mean ± s.e.m.

665

range regardless of light intensity or spike probability (Fig. 3h), indicating that the gamma oscillations evoked by FS activity are a resonant circuit property. In addition, randomly patterned light stimulation of FS cells with frequencies evenly distributed across a broad range evoked a significant increase in LFP power specific to the gamma range ($n = 7$ sites in 4 animals; $P < 0.05$; Supplementary Fig. 7), further indicating that FS-evoked gamma oscillations are an emergent property of the circuit and do not require exclusive drive in the gamma range.

### Natural gamma oscillations require FS activity

To test whether intrinsically occurring gamma oscillations show a similar dependence on FS activity, we gave single light pulses during epochs of natural gamma. We found that brief FS activation shifted the phase of both spontaneously occurring gamma oscillations ($n = 26$ trials, 4 animals; Kruskal–Wallis test with Dunn's post-test; $P < 0.01$; Fig. 3i) and those evoked by midbrain reticular formation stimulation ($n = 18$ trials, 2 animals; $P < 0.05$; Supplementary Fig. 8). Furthermore, light-induced gamma oscillations were largely eliminated by blocking AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) and NMDA (N-methyl-D-aspartate) receptors, despite high levels of evoked FS firing ($n = 4$ sites in 4 animals; $P < 0.01$; Supplementary Fig. 9). These results indicate that induced gamma oscillations depend on rhythmic excitatory synaptic activity, as predicted by computational models of natural gamma oscillations and previous experiments[4,9,11,12,26]. In further agreement, spontaneous RS activity was entrained by 40 Hz FS stimulation, resulting in RS firing during the decay phase of the IPSP and preceding subsequent evoked FS spiking (Supplementary Fig. 10).

### Evoked gamma phase regulates sensory processing

Gamma oscillations are thought to have a functional impact on cortical information processing by synchronizing the output of excitatory neurons[27,28]. This synchrony selects cell assemblies involved in a common task, such as encoding a sensory stimulus, and enhances their impact on downstream targets[27]. The cyclical FS inhibition underlying gamma oscillations is believed to cause this synchrony by rhythmically gating synaptic inputs[27,29]. Synaptic inputs arriving at the peak of inhibition should therefore produce a diminished response, but those arriving at the opposite phase in the gamma cycle should evoke a large response.

To test this hypothesis directly, we stimulated FS cells at 40 Hz with light pulses to establish gamma oscillations, and recorded the responses of RS cells to a single vibrissa deflection presented at one of five phases relative to a single gamma cycle ($n = 20$ cells in 3 animals; Fig. 4a). The timing of vibrissa-induced RS action potentials relative to light-evoked inhibition and the gamma cycle had a significant impact on the amplitude, timing and precision of the sensory-evoked responses of RS cells (Fig. 4b, c). The presence of gamma oscillations significantly decreased the amplitude of the RS sensory response at three phase points, consistent with the enhanced level of overall inhibition in this state ($P < 0.05$; 1-way ANOVA with Dunnett's post-test; Fig. 4d)[28]. Gamma phase also modulated the overall timing of the sensory response ($P < 0.01$; Fig. 4e), with spike latency delayed at phases 1−3 and unaffected at phases 4−5 (ref. 28). The precision of sensory-evoked spikes was significantly enhanced in a gamma-phase-dependent manner ($P < 0.01$; Fig. 4f). Our results indicate that the rhythmic, FS-induced IPSP restricts sensory transmission during its peak, and permits transmission after its decay, leading to a temporal sharpening of cortical sensory responses (Fig. 4g).

Our results provide the first causal demonstration of cortical oscillations induced by cell-type-specific activation. Synchronous FS-PV[+] interneuron activity driven by periodic stimulation of light-activated channels generated gamma oscillations in a cortical network, and these gated sensory processing in a temporally specific manner. These findings also demonstrate a unique application of optogenetic engineering in the *in vivo* brain for the study of discrete neuronal cell
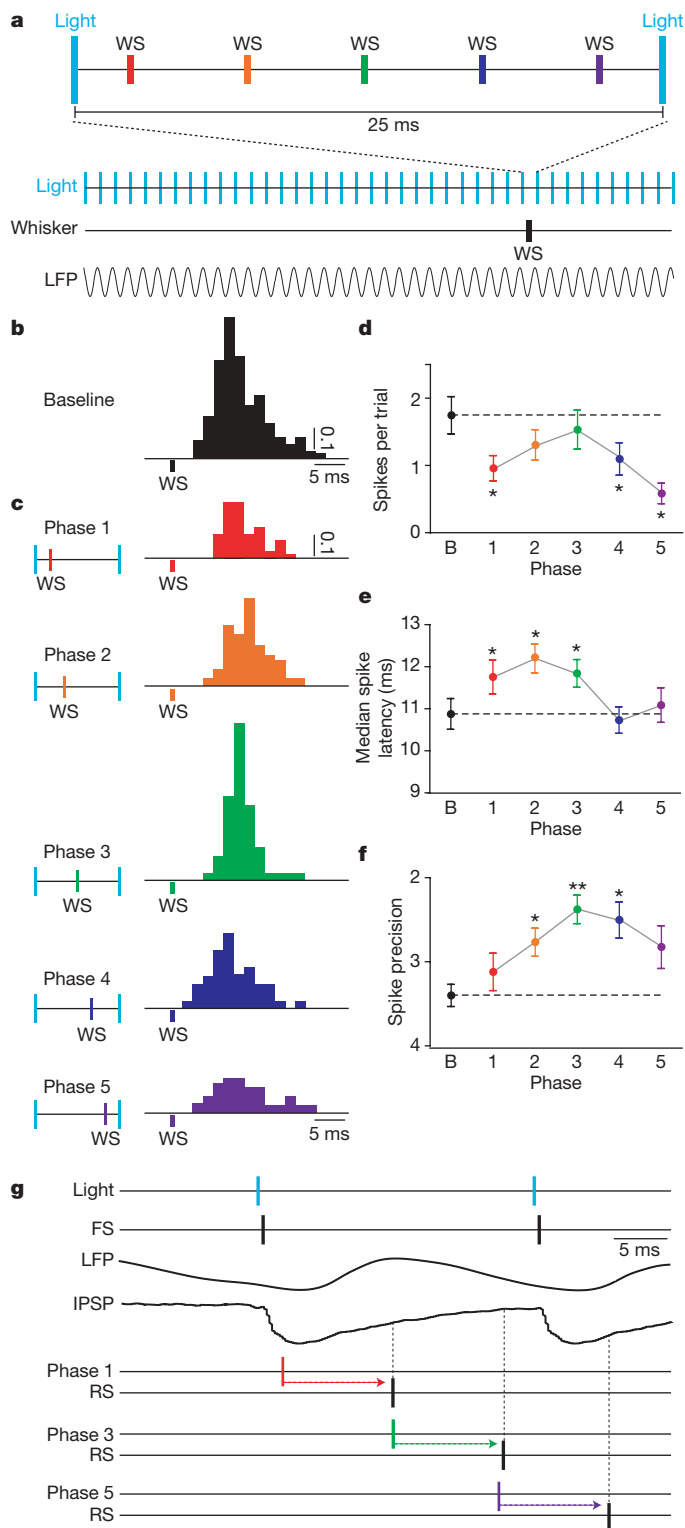


**Figure 4 | Gamma oscillations gate sensory responses of excitatory neurons. a**, In each trial, FS-PV[+] inhibitory interneurons were activated at 40 Hz and a single vibrissa deflection (whisker stimulus, WS) was presented at one of five phases. **b**, Baseline response of one layer 4 RS cell to single vibrissa deflections, shown in units of spikes per trial. **c**, Responses of the same cell when the whisker was deflected at each of five temporal phases relative to the induced gamma oscillation. **d**, Average spikes evoked per trial under each condition. Dotted line indicates baseline responses. **e**, Timing of the RS spike response, measured as median spike latency. **f**, Spike precision of the RS responses. **g**, Schematic model of the gating of sensory responses by gamma oscillations. IPSP and LFP examples are averaged data traces. *$P < 0.05$, **$P < 0.01$; error bars, mean ± s.e.m.

types under active network conditions. Future use of these techniques will allow direct testing of the impact of brain states on information processing in the behaving animal[30], and potentially the rescue of functional states in models of brain disease[31–33].

## METHODS SUMMARY

AAV DIO *ChR2−mCherry* was injected into barrel cortex in adult PV−Cre[18] and CW2 (ref. 24) mice. Six days after injection, a subset of mice was perfused and tissue sectioned for immunohistochemistry to confirm the location, transduction efficiency and cell type specificity of ChR2 expression. Spread and immunohistochemical analysis of ChR2−mCherry expression was scored by hand through examination of every $30\,\mu m$ coronal section for the presence of mCherry fluorescence. For electrophysiology, mice were anesthetized with isoflurane, and extracellular single-unit and LFP recordings were made in layers 2/3 and 4 of barrel cortex with tetrodes or stereotrodes. Intracellular recordings were made in the whole-cell configuration. Light pulses were given by means of an optical fibre positioned at the cortical surface directly above the recording site. For experiments using a broad range of light stimulation frequencies, we stimulated in bouts of 3 s at each frequency in a random order. For each stimulation frequency, we measured relative power in an 8 Hz band centred on that frequency. Relative power was calculated by measuring the ratio of power within the band of interest to total power in the power spectrum. To illustrate the impact of stimulation on the LFP, we also measured the ratio of power in a band with light stimulation to power in the band under baseline conditions. Vibrissae were stimulated by computer-controlled movements of piezoelectric wafers and consisted of single high-velocity deflections in the dorsal direction. For gamma-phase experiments, we gave a series of trials each consisting of a 1-s series of 1-ms light pulses at 40 Hz, with a single whisker deflection after the thirtieth light pulse. The precise timing of the whisker deflection relative to the light pulses was varied across five phase points. All numbers are given as mean ± s.e.m., except where otherwise noted.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Berger, H. On the electroencephalogram of man. *Electroencephalogr. Clin. Neurophysiol.* **28** (Suppl.) 37–74 (1969).
2. Steriade, M. Grouping of brain rhythms in corticothalamic systems. *Neuroscience* **137,** 1087–1106 (2006).
3. Traub, R. D., Whittington, M. A., Stanford, I. M. & Jefferys, J. G. A mechanism for generation of long-range synchronous fast oscillations in the cortex. *Nature* **383,** 621–624 (1996).
4. Traub, R. D., Jefferys, J. G. & Whittington, M. A. Simulation of gamma rhythms in networks of interneurons and pyramidal cells. *J. Comput. Neurosci.* **4,** 141–150 (1997).
5. Whittington, M. A., Traub, R. D. & Jefferys, J. G. Synchronized oscillations in interneuron networks driven by metabotropic glutamate receptor activation. *Nature* **373,** 612–615 (1995).
6. Whittington, M. A., Faulkner, H. J., Doheny, H. C. & Traub, R. D. Neuronal fast oscillations as a target site for psychoactive drugs. *Pharmacol. Ther.* **86,** 171–190 (2000).
7. Deans, M. R., Gibson, J. R., Sellitto, C., Connors, B. W. & Paul, D. L. Synchronous activity of inhibitory networks in neocortex requires electrical synapses containing connexin36. *Neuron* **31,** 477–485 (2001).
8. Galarreta, M. & Hestrin, S. A network of fast-spiking cells in the neocortex connected by electrical synapses. *Nature* **402,** 72–75 (1999).
9. Hasenstaub, A. *et al.* Inhibitory postsynaptic potentials carry synchronized frequency information in active cortical networks. *Neuron* **47,** 423–435 (2005).
10. Wang, X. J. & Buzsaki, G. Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model. *J. Neurosci.* **16,** 6402–6413 (1996).
11. Borgers, C., Epstein, S. & Kopell, N. J. Background gamma rhythmicity and attention in cortical local circuits: a computational study. *Proc. Natl Acad. Sci. USA* **102,** 7002–7007 (2005).
12. Whittington, M. A., Traub, R. D., Faulkner, H. J., Stanford, I. M. & Jefferys, J. G. Recurrent excitatory postsynaptic potentials induced by synchronized fast cortical oscillations. *Proc. Natl Acad. Sci. USA* **94,** 12198–12203 (1997).
13. Gray, C. M. & Singer, W. Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl Acad. Sci. USA* **86,** 1698–1702 (1989).
14. Fries, P., Reynolds, J. H., Rorie, A. E. & Desimone, R. Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* **291,** 1560–1563 (2001).
15. Fries, P., Nikolic, D. & Singer, W. The gamma cycle. *Trends Neurosci.* **30,** 309–316 (2007).
16. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neurosci.* **8,** 1263–1268 (2005).
17. Deisseroth, K. *et al.* Next-generation optical technologies for illuminating genetically targeted brain circuits. *J. Neurosci.* **26,** 10380–10386 (2006).
18. Hippenmeyer, S. *et al.* A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biol.* **3,** e159 (2005).
19. Kuhlman, S. J. & Huang, Z. J. High-resolution labeling and functional manipulation of specific neuron types in mouse brain by Cre-activated viral gene expression. *PLoS ONE* **3,** e2005 (2008).
20. Ascoli, G. A. *et al.* Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature Rev. Neurosci.* **9,** 557–568 (2008).
21. Ren, J. Q., Aika, Y., Heizmann, C. W. & Kosaka, T. Quantitative analysis of neurons and glial cells in the rat somatosensory cortex, with special reference to GABAergic neurons and parvalbumin-containing neurons. *Exp. Brain Res.* **92,** 1–14 (1992).
22. Markram, H. *et al.* Interneurons of the neocortical inhibitory system. *Nature Rev. Neurosci.* **5,** 793–807 (2004).
23. Cauli, B. *et al.* Molecular and physiological diversity of cortical nonpyramidal cells. *J. Neurosci.* **17,** 3894–3906 (1997).
24. Zeng, H. *et al.* Forebrain-specific calcineurin knockout selectively impairs bidirectional synaptic plasticity and working/episodic-like memory. *Cell* **107,** 617–629 (2001).
25. Hubbard, J. I., Llinas, R. & Quastel, D. M. J. *Electrophysiological Analysis of Synaptic Transmission* (The Camelot Press Ltd, 1969).
26. Borgers, C. & Kopell, N. Effects of noisy drive on rhythms in networks of excitatory and inhibitory neurons. *Neural Comput.* **17,** 557–608 (2005).
27. Engel, A. K. & Singer, W. Temporal binding and the neural correlates of sensory awareness. *Trends Cogn. Sci.* **5,** 16–25 (2001).
28. Fries, P., Neuenschwander, S., Engel, A. K., Goebel, R. & Singer, W. Rapid feature selective neuronal synchronization through correlated latency shifting. *Nature Neurosci.* **4,** 194–200 (2001).
29. Burchell, T. R., Faulkner, H. J. & Whittington, M. A. Gamma frequency oscillations gate temporally coded afferent inputs in the rat hippocampal slice. *Neurosci. Lett.* **255,** 151–154 (1998).
30. Huber, D. *et al.* Sparse optical microstimulation in barrel cortex drives learned behaviour in freely moving mice. *Nature* **451,** 61–64 (2008).
31. Orekhova, E. V. *et al.* Excess of high frequency electroencephalogram oscillations in boys with autism. *Biol. Psychiatry* **62,** 1022–1029 (2007).
32. Spencer, K. M., Niznikiewicz, M. A., Shenton, M. E. & McCarley, R. W. Sensory-evoked gamma oscillations in chronic schizophrenia. *Biol. Psychiatry* **63,** 744–747 (2008).
33. Uhlhaas, P. J., Haenschel, C., Nikolic, D. & Singer, W. The role of oscillations and synchrony in cortical networks and their putative relevance for the pathophysiology of schizophrenia. *Schizophr. Bull.* **34,** 927–943 (2008).

**Author Contributions** J.A.C., M.C., K.M., L.-H.T. and C.I.M. designed the experiments. F.Z. and K.D. designed and cloned the AAV DIO ChR2−mCherry vector. M.C. and K.M. characterized the virus *in vitro* and *in vivo* and injected the animals. M.C. performed histological statistical analyses. J.A.C. performed and analysed the extracellular recordings. U.K. and J.A.C. performed the intracellular recordings. U.K. analysed the intracellular data. J.A.C., M.C., K.M., U.K., L.-H.T. and C.I.M. wrote the manuscript.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.I.M. (cim@mit.edu), L.-H.T (lhtsai@mit.edu) or K.D. (deissero@stanford.edu).

*nature*

# METHODS

**Animals.** All procedures were conducted in accordance with the National Institutes of Health guidelines and with the approval of the Committee on Animal Care at MIT. PV$-$Cre ($n = 21$) and CW2 ($n = 7$) mice were $6-12$ weeks old at the time of viral injections. Electrophysiological recordings and immuno-histochemical analyses were performed $1-3$ weeks after viral injections.

**AAV vectors.** ChR2 fused to the fluorescent protein mCherry was cloned in antisense direction into pAAV-MCS (Stratagene) to create AAV DIO *ChR2$-$mCherry* (Fig. 1a and Supplementary Fig. 1; for vector outline and sequence see www.optogenetics.org). *ChR2$-$mCherry* was flanked by a pair of canonical *loxP* sites and a pair of mutated *lox2272* sites. A woodchuck hepatatis B virus post-transcriptional element was placed in sense direction $5'$ of the poly(A). Adeno-associated viral particles of serotype 2 were produced by the Vector Core Facility at The University of North Carolina at Chapel Hill.

**Virus injections.** Adult PV$-$Cre[18] or CW2 (ref. 24) mice were anesthetized with an intraperitoneal injection of a mixture of ketamine ($1.1 \, \text{mg kg}^{-1}$) and xylazine ($0.16 \, \text{mg kg}^{-1}$). A small craniotomy was made 1.5 mm posterior to bregma and 3.0 mm lateral to the midline. Virus was delivered through a small durotomy by a glass micropipette attached to a Quintessential Stereotaxic Injector (Stoelting). The glass micropipette was lowered to 0.4 mm below the cortical surface. A bolus of 0.5 µl of virus (AAV DIO *ChR2$-$mCherry*; $2 \times 10^{12}$ viral molecules per ml) was injected into barrel cortex at $0.1 \, \mu\text{l min}^{-1}$. The pipette was then retracted to a depth of 250 µm below the surface and an additional 0.5 µl virus was injected at the same rate. The pipette was held in place for 5 min after the injection before being retracted from the brain. The scalp incision was sutured, and post-injection analgesics were given to aid recovery ($0.1 \, \text{mg kg}^{-1}$ Buprenex).

**Immunohistochemistry.** Mice were transcardially perfused with 100 mM PBS followed by 4% formaldehyde in PBS, and brains were post-fixed for 18 h at 4 °C. Free-floating sections (30 µm) were cut using a vibratome (Leica VT100) and incubated with blocking solution (10% donkey serum in PBS with 0.3% Triton-X 100) for 1 h at room temperature (20 °C) and then incubated at room temperature overnight with primary antibody diluted in blocking solution. The following primary antibodies were used: NeuN (Chemicon; 1:1,000), parvalbumin PVG-214 (Swant; 1:2,000), GABA (Sigma; 1:4,000) and CamKII (Epitomics; 1:500). After washing, antibody staining was revealed using species-specific fluorophore-conjugated secondary antibodies (Cy5 from Jackson Laboratories, Alexa 488 from Molecular Probes). GABA was detected with biotinylated secondary antibodies (Jackson Laboratories) and revealed using a combination of ABC kit (Vector Laboratories) and TSA fluorescent amplification kit (Perkin-Elmer). Sections were mounted on glass slides with Vectashield (Vector Laboratories) and coverslipped.

**Quantification.** Spread and labelling efficiency were scored by hand by examination of every 30 µm coronal section ($n = 3$ animals per genotype) for the presence of mCherry fluorescence using a Zeiss LSM510 confocal microscope. For quantification of co-labelling of ChR2$-$mCherry and PV ($n = 4$ animals per genotype) confocal images were acquired and individual cells were identified independently for each of the two fluorescent channels. Scans from each channel were collected in multi-track mode to avoid cross-talk between channels.

**Electrophysiology.** Mice were anesthetized with isoflurane and held in place with a head post cemented to the skull. All incisions were infiltrated with lido-caine. A small craniotomy was made over barrel cortex approximately 200 µm anterior to the virus injection site. Extracellular single-unit and LFP recordings were made with tetrodes or stereotrodes. Intracellular recordings were conducted by whole-cell *in vivo* recording in current clamp mode. Stimulus control and data acquisition was performed using software custom-written in LabView (National Instruments) and Matlab (The Mathworks). Further electrophysiology methods and a description of the reversal potential calculation are given in Supplementary Methods.

Light stimulation was generated by a 473 nm laser (Shanghai Dream Lasers) controlled by a Grass stimulator (Grass Technologies) or computer. Light pulses were given via a 200-µm diameter, unjacketed optical fibre (Ocean Optics) positioned at the cortical surface $75-200 \, \mu\text{m}$ from the recording electrodes. For experiments using the broad range of light-stimulation frequencies (8, 16, 24, 32, 40, 48, 80, 100 and 200 Hz), we stimulated in bouts of 3 s of 1-ms pulses at $46 \, \text{mW mm}^{-2}$ at each frequency in a random order. In a subset of these experiments, we stimulated at 31, 46 and $68 \, \text{mW mm}^{-2}$.

Vibrissae were stimulated by computer-controlled movements of piezoelectric wafers (Piezo Systems). Vibrissa stimulations were single high-velocity deflections in the dorsal and then in the ventral direction (~6 ms duration). In most cases, adjacent vibrissae that yielded indistinguishable amplitude responses during hand mapping were deflected simultaneously. Vibrissa stimulations evoked layer 4 RS spike responses with an onset latency of $9.1 \pm 0.08 \, \text{ms}$. For RS cell response suppression experiments, light pulses were given on randomly interleaved trials. For gamma-phase experiments, we gave a series of trials each consisting of a 1-s series of 1-ms light pulses at 40 Hz, with a single whisker deflection after the thirtieth light pulse. The precise timing of the whisker deflection relative to the light pulses was varied across five phase points. Each of the five phase points was included in a random order across a minimum of 250 total trials.

Unit and LFP analysis used software custom-written in Igor Pro (Wavemetrics). For each stimulation frequency, we measured the relative power in an 8-Hz band centred on that frequency. For each recording site, we measured power from $5-10$ LFP traces under each condition. Example power spectra are averages of the power spectra from $5-10$ traces of unfiltered LFPs from individual experiments. Relative power was calculated by measuring the ratio of power within the band of interest to total power in the power spectrum of the unfiltered LFP. We also measured the power ratio: $P_{\text{light}}/P_{\text{baseline}}$, where $P_{\text{light}}$ is the relative power in a frequency band in the presence of light stimulation and $P_{\text{baseline}}$ is the power in that band in the absence of light stimulation. All numbers are given as mean $\pm$ s.e.m., except where otherwise noted.

# ARTICLES

# Chaperonin overexpression promotes genetic variation and enzyme evolution

Nobuhiko Tokuriki[1] & Dan S. Tawfik[1]

**Most protein mutations, and mutations that alter protein functions in particular, undermine stability and are therefore deleterious. Chaperones, or heat-shock proteins, are often implicated in buffering mutations, and could thus facilitate the acquisition of neutral genetic diversity and the rate of adaptation. We examined the ability of the *Escherichia coli* GroEL/GroES chaperonins to buffer destabilizing and adaptive mutations. Here we show that mutational drifts performed *in vitro* with four different enzymes indicated that GroEL/GroES overexpression doubled the number of accumulating mutations, and promoted the folding of enzyme variants carrying mutations in the protein core and/or mutations with higher destabilizing effects (destabilization energies of $>3.5$ kcal mol$^{-1}$, on average, versus $\sim$1 kcal mol$^{-1}$ in the absence of GroEL/GroES). The divergence of modified enzymatic specificity occurred much faster under GroEL/GroES overexpression, in terms of the number of adapted variants ($\geq$2-fold) and their improved specificity and activity ($\geq$10-fold). These results indicate that protein stability is a major constraint in protein evolution, and buffering mechanisms such as chaperonins are key in alleviating this constraint.**

Proteins are thought to evolve under strong stability constraints. Their relatively small margin of stability prevents the acquisition of a large fraction of mutations, and thus severely limits the rate of evolution[1–3]. Mutations that alter protein functions are particularly destabilizing[4,5]. Thus, protein variants that carry mutations with the potential to endow new, or improved, functions are likely to be purged despite, or even because of, their innovative potential[1]. Mechanisms that buffer the destabilizing effects of mutations therefore have a critical role in maintaining higher genetic diversity and expediting the rate of adaptation[1,6].

Chaperones, or heat-shock proteins (HSPs), are essential to eukaryotic and prokaryotic cells as they assist the folding, disaggregation, transport and function of proteins. They buffer the effects of environmental changes, as indicated by the marked increase in their levels under various stress conditions[7,8]. Chaperones also comprise a buffering system to genetic perturbations, and thus promote genetic diversity[9,10]. In eukaryotes, partial reduction of HSP90 results in unmasking of heritable phenotypic variation[11–13]. Eukaryotic chaperones may also mediate adaptive evolution by buffering the deleterious effects of mutated genes that mediate new functions[10]. The *Escherichia coli* GroEL/GroES chaperonins were shown to rescue heat-sensitive protein mutants[14], and bacteria facing high mutational loads are fitter under GroEL/GroES overexpression[15,16]. However, chaperone functions are very complex, in not only assisting protein folding but also affecting the development and signal transduction networks of the cell and alleviating the toxic effects of aggregated proteins. Their buffering mechanisms also vary from one organism, and chaperone, to another[9,16–19], and some cases of chaperone-related adaptations do not involve a direct rescue of mutated proteins[17,19].

We explored GroEL, which, with its cofactor GroES, is the most intensely studied and only chaperone in *E. coli* essential under all growth conditions[8]. Indeed, GroEL/GroES interacts with about 250 different *E. coli* proteins, which constitute ~10% of all soluble *E. coli* proteins[20]. We aimed to establish an *in vitro* experimental system for a systematic and quantitative assessment of the capability of this chaperonin to promote the acquisition of genetic diversity by assisting the folding of destabilized enzyme mutants. We were particularly interested in the potential role of GroEL/GroES in rescuing mutants carrying altered or improved enzymatic functions.

## The experimental system

We performed a random mutational drift of several enzymes in the presence of GroEL/GroES overexpression (Fig. 1). Mutagenesis of the target enzyme gene was performed *in vitro* by error-prone polymerase chain reaction (PCR) optimized to give ~1.5 random mutations per gene per generation (Supplementary Table 1). The mutated genes were transformed to *E. coli* for overexpression, in parallel with GroEL/GroES overexpression or in its absence. Colonies carrying individual enzyme variants were grown, the bacteria were lysed and the levels of enzymatic activity were assayed. The enzyme-coding genes were re-cloned after each round of mutagenesis and screening, and subsequently transformed into a fresh sample of *E. coli* (Fig. 1). This procedure eliminated the role of factors related to bacterial fitness and other intergenic effects, and restricted the study to specific mutations within the coding region of the target enzymes and to the effect of GroEL/GroES overexpression on these mutations.

We tested four enzymes: glyceraldehydophosphate dehydrogenase from *E. coli* (GAPDH) and human carbonic anhydrase 2 (CA2), which are known clients of GroEL (proteins bound and refolded by GroEL)[20,21]; a variant of *Pseudomonas* sp. phosphotriesterase (PTE), named PTE-vr (see Methods), which was confirmed to be a GroEL/GroES client by our experiments; and *E. coli* triose phosphate isomerase (TIM), which is not a GroEL/GroES client[20,22]. All four enzymes were overexpressed in a soluble and functional form in *E. coli* in a chaperonin-independent manner: protein levels, indicated by both SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and enzyme activity, were similar under GroEL/GroES overexpression (growth with arabinose) compared to no chaperonin overexpression (growth with glucose).

As expected, mutagenesis led to a substantial decrease in enzymatic activity, and 20–40% of the 360 randomly picked variants in the initial screen showed no measurable activity. Enzyme variants that retained

[1]Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel.
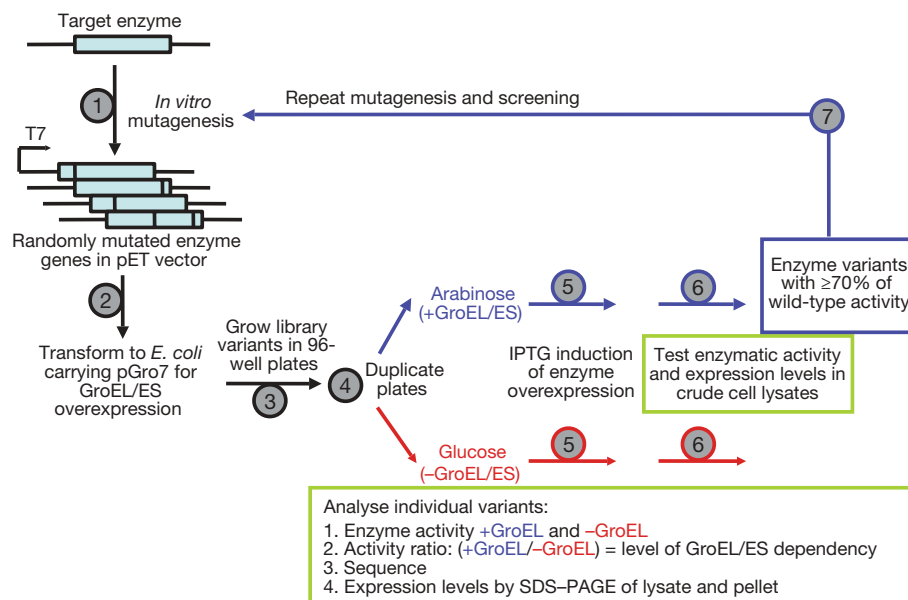
**Figure 1 | The experimental setup.** The target enzymes were subjected to random mutagenesis (step 1). The mutated genes were cloned into a pET expression vector and transformed to *E. coli* carrying the pGro7 plasmid, which conditionally overexpressed GroEL/GroES (step 2). In each round, 360 randomly picked colonies were individually re-grown in liquid media within 96-well plates (step 3). The plates were duplicated. To one set of plates, arabinose was added to induce GroEL/GroES overexpression. Glucose was added to the parallel set to suppress GroEL/GroES expression (step 4). Overexpression of the mutated enzyme variants was subsequently induced by IPTG addition (step 5). The cells were lysed, enzyme activity with and without GroEL/GroES overexpression was recorded (step 6), and representative variants were analysed (green box). To continue the drift, clones showing activity above the chosen threshold under arabinose induction were taken for the next round of mutagenesis and screening (step 7).

sufficient activity (150 individual clones, on average) were taken for the next round of mutagenesis and screening (Fig. 1, step 7). For GAPDH, PTE and TIM, we defined the threshold as ≥70% of the wild-type activity under GroEL/GroES overexpression. The threshold in the case of CA2, which was less tolerant to mutagenesis, was defined as ≥30%. The mutational drift, or 'mutation-accumulation' experiment, therefore occurred under purifying selection to purge deleterious mutations, and the soluble expression and enzymatic function of the drifting enzyme variants was maintained throughout. The overexpression of GroEL/GroES, the high mutation rates, small population sizes and relatively stringent selections used here exaggerate the effects of both GroEL/GroES and mutations. However, these conditions were necessary to enable the monitoring of GroEL/GroES effects for hundreds of different mutations, and thereby provide a clear and quantitative picture of how buffering the destabilizing effects of mutations accelerates neutral and adaptive evolution. Potential implications regarding more realistic evolutionary scenarios are mentioned in the Discussion.

## Divergence of GroEL/GroES-dependent variants

In the first round, 93% of the GADPH variants were active above the defined activity threshold (Fig. 2, upper panel; the dashed vertical line separates all variants above the threshold of ≥70% of wild-type activity). However, when the drifting variants were tested in parallel with no GroEL/GroES overexpression, only 85% were sufficiently active. By the third generation, 60% of variants were sufficiently active with GroEL/GroES, but only 43% in its absence. Overall, ~30% of the drifting variants had chaperonin-dependent activity, whereby the activity observed with and without GroEL/GroES overexpression varied by a factor of between 2 and 30. In the fourth generation, the percentage of active variants actually increased, and the population did not become more GroEL/GroES-dependent. This tendency was observed when the GADPH drift was repeated (Supplementary Fig. 1a), and is probably due to the drift reaching a stage where the stability of the enzyme is compromised to the level that non-synonymous mutations do not readily accumulate (Supplementary Table 1)[23]. The enzymes PTE and CA2 had a similar behaviour. By the third generation, the fraction of variants for which

levels of activity are chaperonin-dependent became 44% and 34%, respectively (Supplementary Table 2 and Supplementary Fig. 1b, c).

A control GADPH drift in the absence of GroEL/GroES overexpression indicated that none of the drifting variants that retained sufficient activity (Fig. 2, lower panel, left of the dashed line) was GroEL/GroES-dependent (ratio of +GroEL/−GroEL of ≤2). Chaperonin-dependent variants were observed among low-activity variants (right of the dashed line), but they gradually disappeared as the drift progressed. Drifts without GroEL/GroES overexpression were not performed with the other enzymes owing to the laborious nature of these experiments. However, the comparison of mutations found in GroEL/GroES-dependent and -independent variants validated the role of GroEL/GroES in rescuing stability-compromised variants in both GADPH and PTE (next section). An additional control experiment was performed with TIM—an enzyme that does not seem to interact with GroEL/GroES[20,22]. The drifting TIM populations showed a relatively small fraction of GroEL/GroES-dependent variants (4–10%; Supplementary Fig. 1d). The results of all drift experiments are summarized in Supplementary Table 2.

Because both the target enzymes and GroEL/GroES were overexpressed (≥10,000 molecules per cell), the effect of the endogenous chaperones was negligible. To support further the direct role of GroEL/GroES in rescuing the drifted enzyme variants, we compared the active enzyme levels for a representative sample of variants in the presence of overexpression of GroEL on its own (with no overexpression of GroES) and of two additional *E. coli* chaperones: DnaJ/DnaK/GrpE and trigger factor. These led to almost no increases in activity under the conditions in which GroEL/GroES overexpression led to 2–17-fold increases (Supplementary Table 3 and Supplementary Fig. 2). Pool-down assays also indicated the interaction of the target enzymes and their mutants with GroEL/GroES (Supplementary Fig. 3).

## Characteristics of GroEL/GroES-dependent variants

Analysis of the type, location and stability effects of mutations revealed notable differences between neutral variants that drifted with GroEL/GroES assistance (≥2-fold difference in activity with and without GroEL/GroES overexpression) and those that did not (a ratio of

**a** Presence of GroEL/GroES



**b** Absence of GroEL/GroES



+GroEL/GroES   −GroEL/GroES   Ratio (+GroEL/−GroEL)
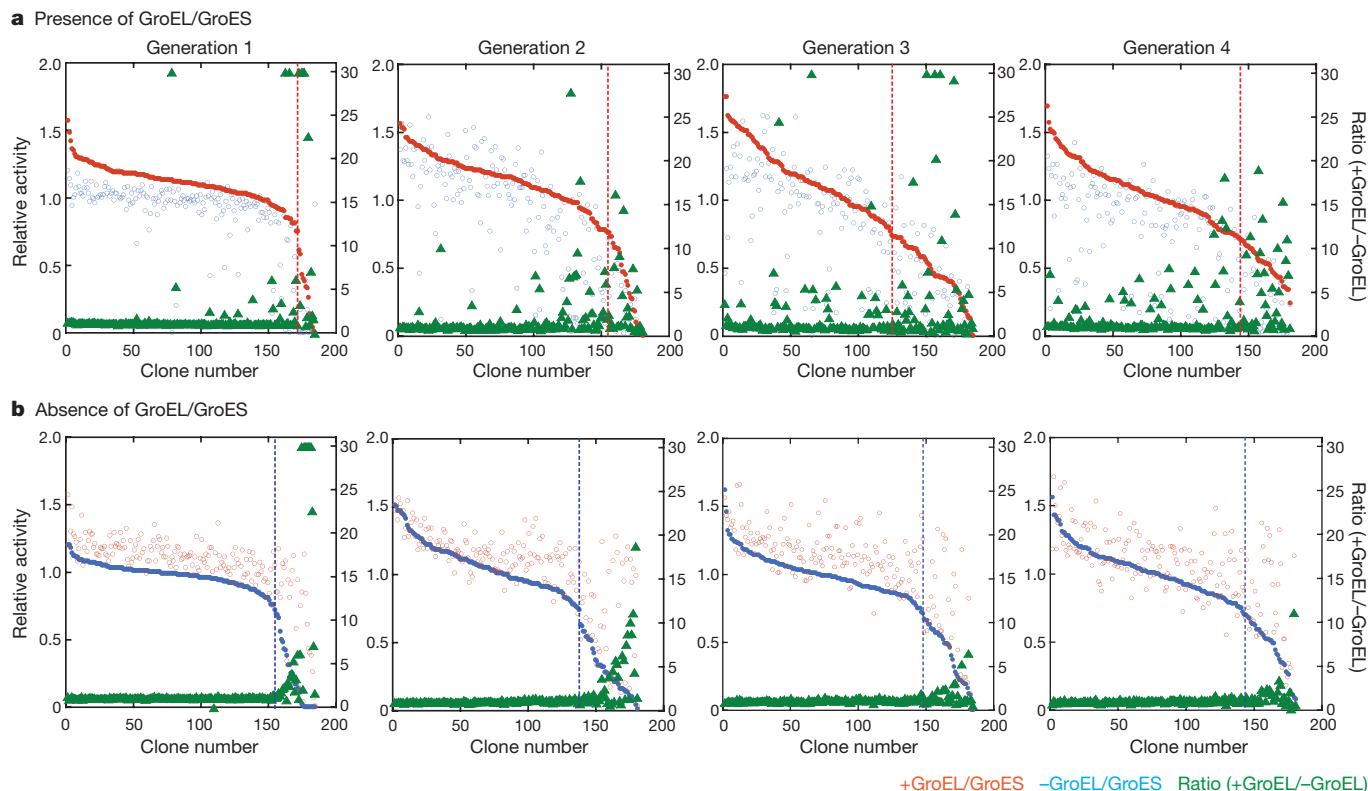
**Figure 2 | GAPDH mutation-accumulation experiments. a,** GADPH was drifted while overexpressing GroEL/GroES. Enzymatic activity was assayed for 180 randomly picked clones, and for each clone activity with GroEL/GroES overexpression relative to wild-type GADPH is presented (red circles; left $y$ axis). Variants exhibiting ≥70% of wild-type activity (left of the dashed vertical line) were taken for the next round of mutagenesis and selection. Each clone was expressed in parallel without GroEL/GroES overexpression (blue open circles; left $y$ axis), and the ratio of activities (+GroEL/−GroEL) is also presented (green triangles; right $y$ axis). **b,** GADPH drifted under the same conditions, yet without GroEL/GroES overexpression. Presented are the activities with and without GroEL/GroES overexpression (red open circles and blue circles, respectively; left $y$ axis) and the GroEL/GroES-dependency (green triangles) for each clone. Variants exhibiting ≥70% of wild-type activity (left of the dashed vertical line) were taken for the next round.

<2). In total, 139 variants and 559 mutations for GAPDH from both the third and fourth generations of the drifts were analysed. The GroEL/GroES-independent variants included variants from the drift in the absence of GroEL/GroES, and variants from the drift under GroEL/GroES overexpression that show no chaperonin dependency (both groups showed similar features). For a start, the GroEL/GroES-independent variants were found to carry less than half the number of amino acid exchanges, or non-synonymous mutations, per gene ($N_a$): an average $N_a$ value of 1.6, on average, versus 3.5 in the GroEL/GroES-dependent variants (all values, and analysis of their statistical significance, are provided in Supplementary Table 4). The fraction of non-synonymous mutations was accordingly lower: $N_a/(N_a + N_s) = 0.46$ versus 0.69 (where $N_s$ is the average number of synonymous mutations, and $N_a + N_s$ is the total number of mutations).

The GroEL/GroES-dependent variants also tolerated a higher fraction of mutations in core residues. Mutations in buried residues are more likely to yield aggregated folding intermediates, and are more destabilizing than exposed surface mutations (average *Escherichia* destabilization energies, $\Delta\Delta G$ values, of $+1.4\,\mathrm{kcal\,mol^{-1}}$, versus $+0.6$)[24]. Consequently, evolutionary rates of core residues are much slower[25]. Indeed, 48% of the mutated residues in GroEL/GroES-dependent variants exhibited relatively low exposed surface areas: ASA values (relative accessible surface area, whereby 0 and 1 correspond to totally buried and exposed residues, respectively) of ≤0.25, compared to 41% in the GroEL/GroES-independent variants. The fraction of highly exposed residues mutated in the absence of GroEL/GroES (ASA values ≥0.6) was concomitantly higher (30% versus 21% with GroEL/GroES overexpression; Supplementary Fig. 4). In agreement with the core–surface distributions, the mutations found in the GroEL/GroES-dependent variants are predicted to be more destabilizing: the average $\Delta\Delta G$ per mutation is $+1.0\,\mathrm{kcal\,mol^{-1}}$,

compared to $+0.6$ for the GroEL/GroES-independent variants. The GroEL/GroES-independent variants also carry threefold more stabilizing mutations: 9% of mutations with $\Delta\Delta G$ of $< -1\,\mathrm{kcal\,mol^{-1}}$, compared with 3% in the GroEL/GroES-dependent variants. Altogether, the GroEL/GroES-dependent variants carried more than twice the number of core mutations, and had a predicted average stability loss per variant of $+3.5\,\mathrm{kcal\,mol^{-1}}$, compared to $+0.9$ for GroEL/GroES-independent variants. The PTE variants from the third generation of the drift (in total 131 variants and 315 mutations) showed a very similar trend (Supplementary Table 4 and Supplementary Fig. 4). There are obviously additional chaperonin effects that relate to folding intermediates, but the FoldX stability-prediction algorithm can only address the native state.

The increase in activity in the presence of GroEL/GroES overexpression is largely due to the ability of the chaperonin to assist the folding of mutants with compromised stability and problematic folding intermediates, thus giving rise to a larger fraction of soluble, active enzyme. This hypothesis was supported by the observed correlation between enzyme activity and the relative occurrence in the soluble fraction versus that in the aggregated fraction (Supplementary Table 5 and Supplementary Fig. 5). As exemplified (Fig. 3a), wild-type GADPH had similar activity with or without GroEL/GroES overexpression, and in both cases >90% of the enzyme was in the soluble fraction. In contrast, drifting GADPH variants that exhibited GroEL/GroES dependency showed a low soluble fraction in the absence of chaperonin overexpression, and a much higher fraction in its presence.

**Divergence of new function**

The early steps in the divergence of new enzymes are presumed to involve mutations that increase a latent, weak promiscuous activity. Such activities may provide an initial selective advantage under changing

**670**
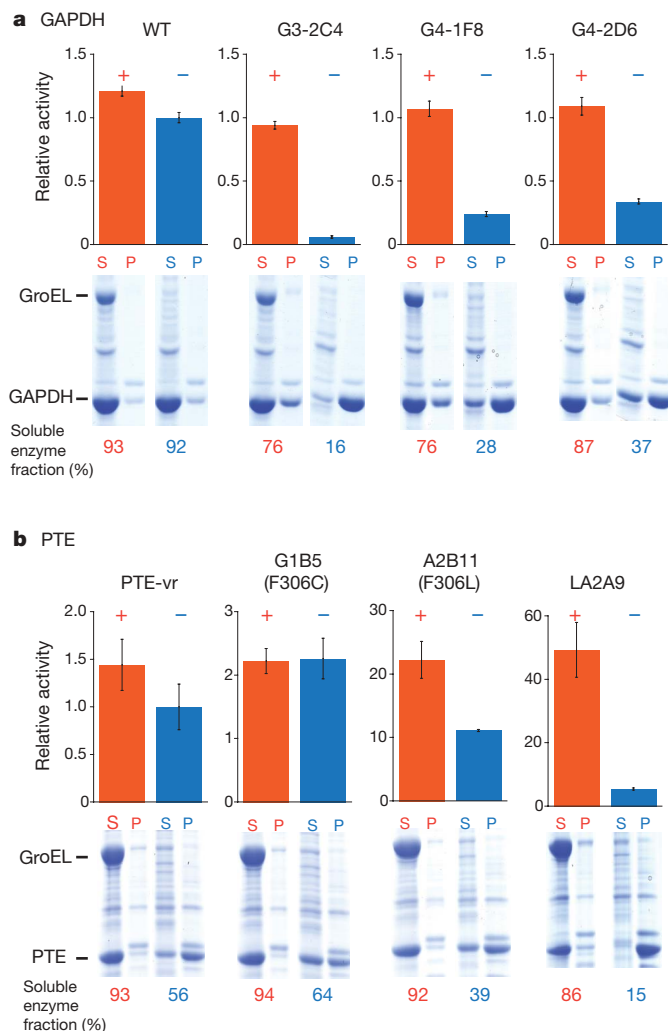
**Figure 3 | Changes in activity and protein levels of enzyme variants expressed with and without GroEL/GroES overexpression. a,** Wild-type GAPDH and its representative neutrally drifting variants. **b,** The starting point for the PTE drift (PTE-vr) and its variants with improved esterase activity. The bars represent relative enzymatic activities in crude lysates with GroEL/GroES overexpression (red) and without it (blue). Error bars represent standard deviations of six independent repeats. The supernatant and insoluble pellet fractions of the above lysates (S and P, respectively) were analysed by SDS–PAGE, and the percentage of enzyme in the soluble fraction was determined by the relative intensities of the supernatant and pellet bands.

environments, and can be subsequently improved by adaptive mutations and become the primary activity of the evolving enzyme[26,27]. However, mutations that endow enzymes with new, or improved, substrate and reaction specificities are, on average, destabilizing and tend to occur in more buried residues than neutral, non-adaptive mutations[4,5]. To examine whether GroEL/GroES may alleviate these stability constraints, we choose PTE—an enzyme that hydrolyses the pesticide paraoxon with high catalytic efficiency ($k_{cat}/K_M$ of $>10^7\,M^{-1}s^{-1}$). We examined the adaptive evolution of PTE to increase its weak, promiscuous esterase activity ($k_{cat}/K_M$ for the target ester substrate 2-naphtylhexanoate is $\sim 10^2\,M^{-1}s^{-1}$). PTE-vr was subjected to random mutagenesis as before, and the resulting library ($\sim 10,000$ transformants) was subjected to a chromogenic screen whereby colonies exhibiting esterase activity turn red.

The number of red colonies in the library expressed with GroEL/GroES (arabinose induction) was far greater than in its absence (agar with glucose; data not shown). The red-coloured colonies were subsequently grown, lysed and tested for enzymatic activity with 2-naphtylhexanoate in 96-well plates. This screen also indicated a significant

difference in favour of the selection with chaperonins, in terms of the number and the fold-improvements in esterase activity (Fig. 4a). Differences in the selectivity of the evolved variants (esterase/paraoxonase rates; Supplementary Table 6) indicated changes in the active site of the mutants, irrespective of changes in their stability and expression.

Almost all the variants showing improved esterase activity carried a mutation of Phe 306 into Cys, Ser or Leu (Supplementary Table 7). Phe306Cys was previously identified in PTE variants evolved towards a similar ester substrate[28]. However, variants selected under GroEL/GroES overexpression, and especially those showing the highest improvements in esterase activity, carried the Phe306Leu mutation that has not been seen in previous attempts to evolve PTE. Two variants carrying only these point mutations were identified: A2B11 (Phe306Leu) from the library with GroEL/GroES overexpression, and G1B5 (Phe306Cys) from the library selected without GroEL/GroES. The Phe306Leu mutation showed higher improvement in esterase activity than Phe306Cys (11.8-fold in $k_{cat}/K_M$, versus 2.6-fold for Phe306Cys), but also had deleterious effects on the folding of PTE that inhibited its identification in the absence of GroEL/GroES overexpression. Indeed, on GroEL/GroES overexpression, the Phe306Leu mutant exhibited a 2.4-fold increase in the levels of soluble protein and a concomitant increase in enzymatic activity. In contrast, the Phe306Cys mutation was barely affected by GroEL/GroES overexpression (Fig. 3b).

The single-mutation variants were subjected to another round of directed evolution with and without GroEL/GroES overexpression, as described previously. The same trends were observed with a larger number of improved variants under GroEL/GroES overexpression, and with higher improvements in both activity (Fig. 4b) and selectivity (Supplementary Table 6). Overall, the evolutionary route under GroEL/GroES produced variants exhibiting up to 44-fold higher esterase activity, and 54-fold higher selectivity, compared to the starting point (PTE-vr). The same route in the absence of GroEL/GroES yielded only 3.4-fold improvements in activity, and 4.2-fold higher selectivity. The kinetic parameters ($k_{cat}$, $K_M$) determined with purified enzymes confirmed the improvements observed in cell lysates (Fig. 4c and Supplementary Table 7).

As observed with neutrally drifting variants, enzymatic activities and expression levels in the soluble versus the aggregated fraction were correlated, indicating that GroEL/GroES overexpression is indeed rescuing the variants evolved in its presence (Supplementary Fig. 5b). Variants selected under GroEL/GroES overexpression showed higher activity under GroEL/GroES overexpression ($\leq 9.1$ versus $\leq 2.4$ for variants selected without GroEL/GroES) and a concomitant increase in the soluble fraction ($\leq 5.8$-fold versus $\leq 2.2$-fold, respectively; Supplementary Table 7). Notably, many variants selected with GroEL/GroES overexpression show little, or no, improvement in activity without GroEL/GroES. For example, in crude lysates variant LA2A9 (diverged from the Phe306Leu mutant) showed twofold improvement in esterase activity with GroEL/GroES overexpression, but twofold decrease without it (Fig. 3b).

As shown with neutral GAPDH and CA2 variants, overexpression of GroEL on its own and of other *E. coli* chaperones (DnaJ/DnaK/GrpE and trigger factor) gave almost no increase in activity with the evolved GroEL/GroES-dependent PTE mutants (Supplementary Table 3). *In vitro* refolding experiments showed that significant refolding was obtained in the presence of GroEL/GroES (10–100%; with wild-type PTE, and with several evolved mutants tested), but no refolding was observed in buffer, or with GroEL on its own (Supplementary Table 7).

## Discussion

Theoretical and experimental studies indicate that seemingly neutral, or hidden, variation has a key role in adaptation towards new environments, at the level of organisms and individual genes[6,26,29,30]. However, the destabilizing effects of mutations can severely limit the acquisition of such neutral variations[2,31]. Over 80% of mutations are deleterious due to their destabilizing effects, which in turn lead to reduced protein
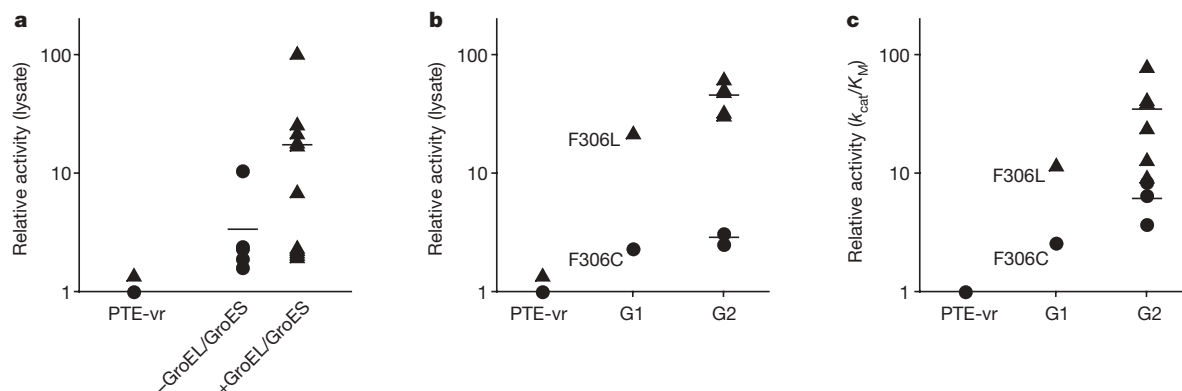
**Figure 4 | Adaptive evolution of *P. diminuta* PTE towards higher esterase activity.** Triangles represent variants evolved and assayed under GroEL/GroES overexpression, and circles represent variants without GroEL/GroES overexpression. Shown are activities relative to the starting point PTE-vr. **a**, Activity levels in crude lysates of variants identified in the first round of mutation and selection. **b**, Activities of second-round variants. Annotated are two single-mutant variants: Phe306Leu (A2B11), which comprised the starting point for the next round under GroEL/GroES overexpression, and Phe306Cys (G1B5) for the selection without it. **c**, $k_{cat}/K_M$ values measured with the purified enzyme variants isolated from these selections (data are provided in Supplementary Table 7).

levels[1,32]. Similarly, most mutations that improve enzymatic activities undermine stability[4,5]. For example, the development of antibiotic resistance in TEM-1 β-lactamase is dependent on the accumulation of a compensatory mutation that makes up for the stability losses associated with mutations that alter the enzyme's substrate specificity[4]. Owing to decreases in the level of soluble, functional enzyme, such mutants may show only small improvement in the cellular rates of enzymatic activity. Decreased stability can also result in insoluble aggregates that reduce fitness[33]. Our results show that rescuing stability-impaired mutants by increasing the fraction of soluble, functional enzyme has marked effects on evolutionary rates.

The experiment also provides a quantitative assessment of the buffering capability of GroEL/GroES. On average, in the absence of GroEL/GroES, the stability of neutral variants seem to be compromised by ~1 kcal mol[−1] (predicted values, assuming additivity of stability effects, are 0.9 and 1.2 kcal mol[−1] for PTE and GADPH, respectively). The overexpression of GroEL/GroES shifted this threshold by about 2.5 kcal mol[−1] to an average level of ~3.6 kcal mol[−1]. A significant enrichment of stabilizing compensatory mutations also occurred in GroEL/GroES-independent variants (Supplementary Fig. 4). Stabilizing mutations are scarce[3,24]. Therefore, evolution, whether neutral or adaptive, that depends on such mutations is far slower. This is indeed reflected in >2-fold lower total number of non-synonymous mutations found in variants drifting without GroEL/GroES (Supplementary Table 4).

GroEL/GroES clients seem to be characterized by low folding propensities and high translation rates (as is the case in our *in vitro* system)[34]. The chaperonin buffered mutants are therefore likely to suffer not only from low stability, but also from lower folding propensities and/or slower folding rates. The buffering capacity of GroEL/GroES is, however, not unlimited. The FoldX predictions indicate that highly destabilizing mutations (>5 kcal mol[−1]) were purged whether GroEL/GroES was overexpressed or not (Supplementary Table 4 and Supplementary Fig. 4). Indeed, many variants that drifted for two, let alone four, generations under GroEL/GroES overexpression appear to be very close to the 5 kcal mol[−1] threshold (Supplementary Table 8), and variants evolved for new function were also severely compromised (for example, LA2A9 described above). Thus, once the extended margin provided by the GroEL/GroES is consumed, the dependency on stabilizing compensatory mutations returns[1,4,23]. The limited buffering capacity is also reflected in the fact that the fraction of mutants that are GroEL/GroES-dependent did not exceed a certain value (for example, ≤35% for the GAPDH drift). This implies that a significant fraction of destabilizing mutations cannot be rescued by GroEL/GroES. The plateau may also be the result of balancing forces such as the appearance of stabilizing compensatory mutations. Indeed, when the GroEL/GroES-dependent PTE variant A2B11 (Phe306Leu) was subjected to evolution in the absence of GroEL/GroES overexpression, a second mutation (Ile274Ser) appeared (variant LG2B6; Supplementary Table 7). Ile274Ser hardly improves catalytic activity (1.1-fold), but it increases stability and lowers GroEL/GroES dependency. This stabilizing mutation may mimic natural processes of reducing chaperone dependency and thus providing 'chaperone space' for other mutants.

The buffering capacity is also protein-dependent, because only a subset of *E. coli* proteins interacts with GroEL/GroES[20,22,34]. Also, drifts we conducted with two known GroEL/GroES clients (green fluorescent protein and TEM-1 β-lactamase data not shown) did not consistently indicate GroEL/GroES-dependency. However, although the wild-type protein does not interact with chaperonins (for example, TIM tested here), its mutants may interact. A number of TIM mutants became GroEL/GroES-dependent, although, relative to the other enzymes tested here, TIM mutants clearly showed a lower tendency to become GroEL/GroES-assisted. The experimental system developed here could be applied to test this correlation in a larger number of proteins.

Our *in vitro* findings have potential implications on natural evolution. Mutation rates in bacteria increase under stress, and so do chaperone levels[35,36]. The combination of increased mutation rates and chaperone overexpression can markedly accelerate the rate of acquisition of new functions, as demonstrated here. However, the balance within living *E. coli* cells, between the capacity to buffer mutations and the rate by which mutations accumulate, is obviously different than the one applied here. Indeed, under stress conditions the available chaperone capacity can also decrease, as shown for HSP90, and the phenotype of previously buffered mutations unravelled[11–13].

Finally, the experimental system described here provides a way of performing directed evolution *in vitro*, whereby, for enzymes that comprise effective chaperonin clients, overexpression of GroEL/GroES can significantly increase the number and fold-improvement of enzyme variants with a new, or improved function—not only at crude cell lysates, but also of the actual kinetic parameters of the evolving enzyme variants (Fig. 4c).

**METHODS SUMMARY**

Wild-type CA2, GAPDH, TIM and engineered PTE-vr were used as the starting points. Gene libraries were created by random mutagenesis using error-prone PCR with the 'wobble' base analogue dPTP, and gene libraries for the evolution of esterase activity of PTE-vr were created with dPTP and 8-oxo-dGTP. The mutated enzyme variants were cloned into pET vector, and transformed to *E. coli*

BL21 (DE3). Agar colonies were individually picked and transferred for growth in 96-well plates. Following arabinose, or glucose, addition, enzyme expression was induced with IPTG. The bacteria were lysed, and the levels of enzymatic activity were assayed in 96-well plates. GAPDH activity was measured by a coupled enzyme assay based on the decrease of absorbance at 340 nm due to NADH oxidation. Paraoxonase activity of PTE was measured by following the increase of absorbance at 405 nm. PTE libraries were screened for esterase activity on agar plates using FAST Red dye that creates a red precipitate with the 2-naphtol product, followed by a spectrophotometeric screen in 96-well plates at 500 nm. The hydrolase activity of CA2 was measured with $p$-nitrophenyl acetate by following the increase of absorbance at 405 nm. TIM activity was determined by a coupled enzyme assay based on the decrease of absorbance at 340 nm due to NADH oxidation. GAPDH and PTE variants were purified over a Strep-tactin Sepharose column (IBA).

**Computations.** The stability effect of mutations ($\Delta\Delta G$ values) found in GAPDH and PTE variants was predicted based on crystal structures (PDB accession codes 1S7C and 1HZY, respectively) using the FoldX algorithm. Surface accessibility (ASA) values of residues were calculated by the web-based program ASA-view (http://gibk26.bse.kyutech.ac.jp/~shandar/netasa/asaview/). A complete list of all mutations and their predicted $\Delta\Delta G$ and ASA values is provided in Supplementary Table 8.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

1. DePristo, M. A., Weinreich, D. M. & Hartl, D. L. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nature Rev. Genet.* **6**, 678–687 (2005).
2. Wilke, C. O., Bloom, J. D., Drummond, D. A. & Raval, A. Predicting the tolerance of proteins to random amino acid substitution. *Biophys. J.* **89**, 3714–3720 (2005).
3. Zeldovich, K. B., Chen, P. & Shakhnovich, E. I. Protein stability imposes limits on organism complexity and speed of molecular evolution. *Proc. Natl Acad. Sci. USA* **104**, 16152–16157 (2007).
4. Wang, X., Minasov, G. & Shoichet, B. K. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J. Mol. Biol.* **320**, 85–95 (2002).
5. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D. S. How protein stability and new functions trade off. *PLOS Comput. Biol.* **4**, e1000002 (2008).
6. Wagner, A. *Robustness and Evolvability in Living Systems.* (Princeton Univ. Press, 2005).
7. Martin, J. & Hartl, F. U. Chaperone-assisted protein folding. *Curr. Opin. Struct. Biol.* **7**, 41–52 (1997).
8. Thirumalai, D. & Lorimer, G. H. Chaperonin-mediated protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 245–269 (2001).
9. Fares, M. A., Moya, A. & Barrio, E. GroEL and the maintenance of bacterial endosymbiosis. *Trends Genet.* **20**, 413–416 (2004).
10. Rutherford, S. L. Between genotype and phenotype: protein chaperones and evolvability. *Nature Rev. Genet.* **4**, 263–274 (2003).
11. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
12. Queitsch, C., Sangster, T. A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624 (2002).
13. Sangster, T. A., Lindquist, S. & Queitsch, C. Under cover: causes, effects and implications of Hsp90-mediated genetic capacitance. *Bioessays* **26**, 348–362 (2004).
14. Van Dyk, T. K., Gatenby, A. A. & LaRossa, R. A. Demonstration by genetic suppression of interaction of GroE products with many proteins. *Nature* **342**, 451–453 (1989).
15. Fares, M. A., Ruiz-González, M. X., Moya, A., Elena, S. F. & Barrio, E. Endosymbiotic bacteria: GroEL buffers against deleterious mutations. *Nature* **417**, 398 (2002).
16. Maisnier-Patin, S. et al. Genomic buffering mitigates the effects of deleterious mutations in bacteria. *Nature Genet.* **37**, 1376–1379 (2005).
17. Cowen, L. E. & Lindquist, S. Hsp90 potentiates the rapid evolution of new traits: drug resistance in diverse fungi. *Science* **309**, 2185–2189 (2005).
18. Sangster, T. A. & Queitsch, C. The HSP90 chaperone complex, an emerging force in plant development and phenotypic plasticity. *Curr. Opin. Plant Biol.* **8**, 86–92 (2005).
19. Bobula, J. et al. Why molecular chaperones buffer mutational damage: a case study with a yeast Hsp40/70 system. *Genetics* **174**, 937–944 (2006).
20. Kerner, M. J. et al. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli. Cell* **122**, 209–220 (2005).
21. Persson, M., Carlsson, U. & Bergenhem, N. C. GroEL reversibly binds to, and causes rapid inactivation of, human carbonic anhydrase II at high temperatures. *Biochim. Biophys. Acta* **1298**, 191–198 (1996).
22. Chapman, E. et al. Global aggregation of newly translated proteins in an *Escherichia coli* strain deficient of the chaperonin GroEL. *Proc. Natl Acad. Sci. USA* **103**, 15800–15805 (2006).
23. Bershtein, S., Golding, K. & Tawfik, D. Intense drifts yield robust and evolvable consensus proteins. *J. Mol. Biol.* **379**, 1029–1044 (2008).
24. Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L. & Tawfik, D. S. The stability effects of protein mutations appear to be universally distributed. *J. Mol. Biol.* **369**, 1318–1332 (2007).
25. Sasidharan, R. & Chothia, C. The selection of acceptable protein mutations. *Proc. Natl Acad. Sci. USA* **104**, 10080–10085 (2007).
26. Amitai, G., Gupta, R. A. & Tawfik, D. S. Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP Journal* **1**, 67–78 (2007).
27. McLoughlin, S. Y. & Copley, S. D. A compromise required by gene sharing enables survival: implications for evolution of new enzyme activities. *Proc. Natl Acad. Sci. USA* **105**, 13497–13502 (2008).
28. Roodveldt, C. & Tawfik, D. S. Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily. *Biochemistry* **44**, 12728–12736 (2005).
29. Le Rouzic, A. & Carlborg, O. Evolutionary potential of hidden genetic variation. *Trends Ecol. Evol.* **23**, 33–37 (2007).
30. Fontana, W. & Schuster, P. Continuity in evolution: on the nature of transitions. *Science* **280**, 1451–1455 (1998).
31. Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N. & Tawfik, D. S. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* **444**, 929–932 (2006).
32. Yue, P., Li, Z. & Moult, J. Loss of protein structure stability as a major causative factor in monogenic disease. *J. Mol. Biol.* **353**, 459–473 (2005).
33. Lindner, A. B., Madden, R., Demarez, A., Stewart, E. J. & Taddei, F. Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proc. Natl Acad. Sci. USA* **105**, 3076–3081 (2008).
34. Noivirt-Brik, O., Unger, R. & Horovitz, A. Low folding propensity and high translation efficiency distinguish *in vivo* substrates of GroEL from other *Escherichia coli* proteins. *Bioinformatics* **23**, 3276–3279 (2007).
35. Radman, M., Matic, I. & Taddei, F. Evolution of evolvability. *Ann. NY Acad. Sci.* **870**, 146–155 (1999).
36. Taddei, F. et al. Role of mutator alleles in adaptive evolution. *Nature* **387**, 700–702 (1997).

**Author Contributions** N.T. and D.S.T. designed the study, analysed the data and wrote the manuscript. N.T. performed the experiments.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.S.T. (tawfik@weizmann.ac.il).

## METHODS

**Gene cloning and library construction.** To identify candidates for the drift experiment, we screened the published lists of chaperone targets[37] for enzymes with chromogenic substrates that can be assayed in crude bacterial lysates. GADPH and CAII met these criteria. Two other candidates (green fluorescent protein and TEM-1 β-lactamase) were explored but the preliminary results were inconclusive. TIM also fulfilled the screening requirement, and served as a control for an enzyme that has been shown not to interact with GroEL/GroES. Finally, PTE (phosphotriesterase from *Pseudomonas diminuta*), for which an array of chromogenic substrates is available[38], was tested and found to be a GroEL/GroES target. Wild-type CA2, GAPDH, TIM and engineered PTE-vr were used as the starting points for the neutral drift. PTE-vr is related to the previously described PTE-S5-1.1, which was evolved for higher *E. coli* expression and 2-naphtyl acetate hydrolase activity[38,39]. The PTE-vr was cloned with NcoI and HindIII sites into a modified pET32-trx[40] for the neutral drift, and into pET-strep for the adaptive evolution experiments. The genes encoding GAPDH and TIM were amplified from the ASKA collection[41] and re-cloned into pET32(b) (Novagen) with NdeI and HindIII sites. For purification and further characterization, GAPDH and its drifted variants were cloned into pET-strep using NcoI and HindIII sites. The CA2-pET construct was described previously[42]. The gene encoding GroEL was amplified from pGro7 plasmid[43] and re-cloned into pBAD33 with NdeI and HindIII sites (pBAD33-GroEL).

Libraries were generated by random mutagenesis with error-prone PCR using the wobble base analogues dPTP and 8-oxo-dGTP (TriLink). The first 20 PCR cycles with primers flanking the cloned genes were performed with BioTaq polymerase (Bioline) in 50-μl reactions using 1 ng of plasmid template, 0.5 μM dPTP for GAPDH, PTE and TIM neutral drifts, 1 μM dPTP for CA2, and 1 μM dPTP and 20 μM 8-oxo-dGTP for PTE's adaptive evolution. The PCR products were treated with DpnI to destroy the template plasmid, purified, and then used as a template (20 ng) for another 15 PCR cycles, this time performed with nesting primers and BioTaq polymerase. The product of this final PCR was restricted with restriction enzymes described above (NEB, 3 h, 37 °C) and ligated into the modified pET as described above. This protocol created with high reproducibility libraries carrying an average of ~1.5 mutations per gene (Supplementary Table 1). This rate provided approximately one amino acid exchange per generation. Lower mutation rate rates would have resulted in screening a large fraction of variants that carry no mutations, and higher rates would result in having multiple mutations appearing simultaneously, and a very low fraction of mutants being viable.

**Activity assays.** Phosphotriesterase activity was assayed by measuring the increase in absorbance at 405 nm with paraoxon (0.5 mM) in 50 mM Tris-HCl, pH 7.5, 0.005% Triton X-100. 2-Naphtylhexanoate (2NH) hydrolysis was assayed by a coupled reaction assay based on the increase of absorbance at 500 nm due to the conjugation of the released 2-naphtol and FAST Red (Sigma) with 0.2 mM 2NH and 0.4 mM FAST Red in 50 mM Tris-HCl, pH 7.5, 0.005% Triton X-100.

GAPDH activity was measured by a coupled enzyme assay based on the decrease of absorbance at 340 nm due to NADH oxidation provided by Sigma. The GAPDH reaction mixture (50 mM Tris-HCl, pH 7.5, 6.7 mM 3-phosphoglyceric acid, 1.1 mM adenosine 5′ triphosphate, 0.12 mM β-nicotinamide adenine dinucleotide, reduced form (β-NADH, Sigma), 3.3 mM L-cysteine, and 8 units per ml 3-phosphoglyceric phosphokinase (Sigma)) was pre-incubated at 30 °C for 1 h, and the reaction was initiated by the addition of the GADPH sample.

TIM activity was determined by a coupled enzyme assay based on the decrease of absorbance at 340 nm due to NADH oxidation[44]. The substrate, glyceraldehyde-3-phosphate, was prepared from DL-glyceraldehyde-3-phosphate diethyl acetal monobarium salt as described by the supplier (Sigma). The TIM reaction mixture contained 1 mM D-glyceraldehyde-3-phosphate, 0.24 mM NADH, and 20 mg ml$^{-1}$ glycerol-3-phosphate dehydrogenase, in 50 mM Tris-HCl (pH 7.5).

The esterase activity of CA2 was assayed by measuring absorbance at 405 nm with 4-nitrophenyl acetate (0.5 mM) in 50 mM Tris-HCl, pH 7.5.

**Coexpression chaperones, target enzymes and activity assays.** *E. coli* BL21 (DE3) cells containing the pGro7 (GroEL/GroES), pBAD33-GroEL (GroEL), pKJE7 (DnaK/DnaJ/GrpE) or pTf16 (trigger factor) plasmid[43] were transformed with the libraries or a variant of GAPDH, PTE, CA2 or TIM, and plated on Luria Broth (LB) agar with 100 μg ml$^{-1}$ ampicillin (amp) and 34 μg ml$^{-1}$ chloramphenicol (cam). Single colonies were picked and grown overnight in 96-deep-well plates at 30 °C with shaking, in LB medium with 100 μg ml$^{-1}$ amp and 34 μg ml$^{-1}$ cam. These overnight cultures were used to inoculate (at 1:20 dilution) within 96-deep-well plates cultures in a fresh LB amp, cam, medium. The medium contained 0.2% (w/v) arabinose or glucose to induce or suppress GroEL/GroES expression, respectively. Cells were grown at 37 °C in the case of GAPDH, CA2 and TIM, or at 30 °C in the case of PTE, with shaking for about two hours, to $D_{600}$ of 0.6–1.0,

and then pelleted and resuspended in LB media pre-incubated at 37 °C or 30 °C with amp and 1 mM IPTG to induce overexpression of the target enzyme. After 1 h incubation at 37 °C in the case of GAPDH, CA2 and TIM, or 30 °C in the case of PTE, the $D_{600}$ of these cultures were measured using a plate reader. Cells were then pelleted and stored at −20 °C overnight. Cells were resuspended in lysis buffer (50 mM Tris-HCl, pH 7.5, 100 μg ml$^{-1}$ lysozyme, 0.5 units per ml benzonase, and 0.1% TritonX-100). The lysates were clarified by centrifugation and assayed for enzymatic activity in 200-μl reactions (20 μl crude lysate was used for assaying CA2; 20 μl of 20-times diluted lysate for GAPDH, or 20 μl of 400-times diluted lysate for PTE; and 20 μl of 400-times diluted lysate for TIM).

**Co-fractionation of GAPDH and PTE with GroEL/GroES.** The enzymes GAPDH or PTE, and mutants thereof, were overexpressed in *E. coli* together with GroEL and His-tagged GroES (or with non-tagged GroES). The cells were lysed with 50 mM Tris-HCl, pH 8.0, 0.01% TritonX-100, 10 mM MgCl₂, 25 units per ml benzonase, 10 mM glucose and 20 units per ml hexokinase. ADP was added immediately after lysis, and the clarified lysate was purified by IMAC affinity chromatography. After 4 washes with 10-times column volume of 20 mM imidazole in Tris pH 8, 500 mM NaCl, 10 mM MgCl₂, 10 mM KCl and 1 mM ADP, the complex was eluted by 500 mM imidazole in Tris, pH 8, and analysed by SDS–PAGE.

**Random mutational drift.** Libraries were transformed into *E. coli* BL21 (DE3) cells carrying the pGro7 plasmid. About 360 transformants were picked into liquid medium in 96-deep-well plates, and GroEL/GroES overexpression was induced with arabinose as described above. At this library size the chances of haphazard fixation are rather low (1/360 < 1%) so the library could maintain high polymorphism, and the sampling is therefore statistically significant as indicated, for example, in the analysis of the mutations. The enzymatic activities were assayed in the crude lysates, and 180 clones showing the detectable activity (≥2-fold above background rates) with GroEL/GroES induction were randomly picked. These 180 variants were then screened in parallel with, and without, overexpression of GroEL/GroES, and the levels of enzymatic activity were determined. The wild type was used as reference in two replica wells. For the subsequent round of neutral drift, variants with the threshold enzymatic activity defined as neutral were picked (≥70% of the wild-type activity of GAPDH, PTE-vr and TIM, and ≥30% of CA2). On average, 150 neutral variants were taken for the next round. Their overnight cultures were combined, and the plasmid DNA purified and used as the template for the next round of mutagenesis by error-prone PCR.

**Directed evolution.** Gene libraries derived from PTE-vr were transformed into *E. coli* BL21 (DE3) cells carry the pGro7 plasmids, and plated on LB-amp-cam agar plates containing 0.2% arabinose (for overexpression of GroEL/GroES) or 0.2% glucose (for suppression of GroEL/GroES). About 10⁴ colonies were screened for each set (arabinose or glucose) using FAST Red[38,42]. From each set, about 280 colonies showing the fastest development of red colour were picked from replica plates, grown in 96-well plates, and assayed for 2NH hydrolysis. Thirty variants that showed the highest 2NH hydrolase activity were picked, and the plasmids were purified and retransformed into fresh *E. coli* BL21 (DE3)-pGro7 competent cells, and then their activities with 2NH and paraoxon were tested in triplicate.

**Enzyme purification and kinetics.** The evolved PTE and GAPDH variants were purified using Strep-tactin affinity columns (IBA) according to the manufacturer's protocol. Initial rates were determined as described above using paraoxon (0.5 to 1,000 μM) and 2NH (2 to 2,000 μM). The kinetics parameters $K_M$ and $K_{cat}$ were determined by fitting the initial rates to the Michaelis–Menten model using KaleidaGraph (Synergy Software). GroEL and GroES were purified as described previously[45].

**Protein refolding of PTE variants.** 10 μM of purified PTE variants was denatured in 30 mM HCl for 1 h at 25 °C and diluted 100-fold (final 0.1 μM PTE) into 50 mM Tris-HCl, pH 7.5, 10 mM KCl, 1 mM MgCl₂ and 1 mM ATP in the absence or presence of 0.2 μM GroEL or GroEL/GroES overnight at 25 °C. The enzymatic activity was measured to determine the fraction of refolded active enzyme.

**Computation of stability effects ($\Delta\Delta G$) by FoldX[46,47].** We applied the published four-step procedure: first, the structures of GAPDH (PDB 1S7C) and PTE (PDB 1HZY) were optimized; second, mutant three-dimensional structures were created by introducing one mutation at a time, and locally optimizing the structure; third, the energy functions of the mutants were computed and compared with the wild-type function; fourth, the actual free energy values were calculated based on the FoldX correlation function between experimental and calculated data ($\Delta\Delta G^{Experimental} = (\Delta\Delta G^{FoldX} + 0.078)/1.14$ (ref. 48).

37. Kerner, M. J. *et al.* Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122**, 209–220 (2005).
38. Roodveldt, C. & Tawfik, D. S. Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily. *Biochemistry* **44**, 12728–12736 (2005).

39. Roodveldt, C. & Tawfik, D. S. Directed evolution of phosphotriesterase from *Pseudomonas diminuta* for heterologous expression in *Escherichia coli* results in stabilization of the metal-free state. *Protein Eng. Des. Sel.* **18**, 51–58 (2005).

40. Aharoni, A. *et al.* Directed evolution of mammalian paraoxonases PON1 and PON3 for bacterial expression and catalytic specialization. *Proc. Natl Acad. Sci. USA* **101**, 482–487 (2004).

41. Kitagawa, M. *et al.* Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res.* **12**, 291–299 (2005).

42. Gould, S. M. & Tawfik, D. S. Directed evolution of the promiscuous esterase activity of carbonic anhydrase II. *Biochemistry* **44**, 5444–5452 (2005).

43. Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H. & Yura, T. Chaperone coexpression plasmids: differential and synergistic roles of DnaK–DnaJ–GrpE and GroEL–GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli. Appl. Environ. Microbiol.* **64**, 1694–1699 (1998).

44. Misset, O. & Opperdoes, F. R. Simultaneous purification of hexokinase, class-I fructose-bisphosphate aldolase, triosephosphate isomerase and phosphoglycerate kinase from *Trypanosoma brucei. Eur. J. Biochem.* **144**, 475–483 (1984).

45. Yifrach, O. & Horovitz, A. Transient kinetic analysis of adenosine 5′-triphosphate binding-induced conformational changes in the allosteric chaperonin GroEL. *Biochemistry* **37**, 7083–7088 (1998).

46. Guerois, R., Nielsen, J. E. & Serrano, L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* **320**, 369–387 (2002).

47. Schymkowitz, J. *et al.* The FoldX web server: an online force field. *Nucleic Acids Res.* **33**, W382–W388 (2005).

48. Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L. & Tawfik, D. S. The stability effects of protein mutations appear to be universally distributed. *J. Mol. Biol.* **369**, 1318–1332 (2007).

# nature

# LETTERS

# A low-energy core-collapse supernova without a hydrogen envelope

S. Valenti[1], A. Pastorello[1], E. Cappellaro[2], S. Benetti[2], P. A. Mazzali[2,3], J. Manteca[4], S. Taubenberger[3], N. Elias-Rosa[5], R. Ferrando[6], A. Harutyunyan[2,7], V. P. Hentunen[8,9], M. Nissinen[8], E. Pian[10], M. Turatto[11], L. Zampieri[2] & S. J. Smartt[1]

**The final fate of massive stars depends on many factors. Theory suggests that some with initial masses greater than 25 to 30 solar masses end up as Wolf–Rayet stars, which are deficient in hydrogen in their outer layers because of mass loss through strong stellar winds. The most massive of these stars have cores which may form a black hole and theory predicts that the resulting explosion of some of them produces ejecta of low kinetic energy, a faint optical luminosity and a small mass fraction of radioactive nickel[1–3]. An alternative origin for low-energy supernovae is the collapse of the oxygen–neon core of a star of 7–9 solar masses[4,5]. No weak, hydrogen-deficient, core-collapse supernovae have hitherto been seen. Here we report that SN 2008ha is a faint hydrogen-poor supernova. We propose that other similar events have been observed but have been misclassified as peculiar thermonuclear supernovae (sometimes labelled SN 2002cx-like events[6]). This discovery could link these faint supernovae to some long-duration γ-ray bursts, because extremely faint, hydrogen-stripped core-collapse supernovae have been proposed to produce such long γ-ray bursts, the afterglows of which do not show evidence of associated supernovae[7–9].**

SN 2008ha was discovered on 7 Nov. 2008 4.1 h UT in the late-type galaxy UGC 12682, at an unfiltered magnitude of 18.8 (ref. 10). A spectrum obtained on 18.18 Nov. 2008 UT (ref. 11) was similar to that of the peculiar supernova SN 2002cx at 10 days past maximum light. The narrow P-Cygni lines are indicative of slowly expanding ejecta, suggesting an outflow velocity 3,000 km s$^{-1}$ slower than that of SN 2002cx. Our spectra show narrow spectral lines confirming the low velocities of the ejected material (2,300 km s$^{-1}$) and no hydrogen signature (see Fig. 1 and Supplementary Information).

The spectral evolution is very fast, and [Ca II] emission at wavelengths $\lambda = 7,291$–7,324 Å was detected one month after the explosion (usually it is visible in core-collapse supernovae only after two to three months). The ejecta velocity slowly decreases with time, reaching ∼1,500 km s$^{-1}$ in the last spectrum. Such low ejecta velocities (∼1,000–1,500 km s$^{-1}$) have been observed in a group of low-luminosity hydrogen-rich core-collapse supernovae[12] (faint type IIP supernovae) with extremely narrow P-Cygni spectral lines, but never in thermonuclear supernovae (type Ia supernovae). These objects are believed to be weak explosions (a few times $10^{50}$ erg) of massive stars, ejecting only $10^{-3}$ solar masses of radioactive material[12].

The photospheric spectrum of SN 2008ha is remarkably similar to that of SN 2005cs (see Fig. 2), apart from the absence of hydrogen lines and the strength of the O I $\lambda = 7,774$ Å feature, which is more pronounced in the spectrum of SN 2008ha than in that of SN 2005cs. The

characteristic type Ia supernova lines (Si II $\lambda = 6,347$–6,371 Å and S II $\lambda = 5,640$ Å), clearly visible in the spectrum of SN 1991T, are distinctly weak or absent in that of SN 2008ha.

Although at early phases SN 2008ha and SN 2005cs are strikingly similar, we know that photospheric spectra alone can be equivocal in discerning between thermonuclear and core-collapse explosions. The existence of prominent Si II and S II lines (combined with a lack of He I and hydrogen features) generally suggests the thermonuclear explosion mechanism. However, Si II lines are also detected in some stripped-envelope core-collapse supernovae (for example, SN 2007gr[13]). To complicate matters, the presence of nearby Fe II, Ti II and Cr II lines blended by high photospheric velocities make the identification of weak Si II rather uncertain and may cause misclassifications.

Additional constraints on the explosion mechanism can be derived from the study of late-time spectra, when the ejecta are optically thin and it becomes easier to probe the nature of the innermost layers. Because of the extremely fast spectral evolution of SN 2008ha, the spectrum at +65 days already shares a remarkable similarity with the nebular spectra of SN 2005cs except for the hydrogen Balmer line (Hα $\lambda = 6,562$ Å; Fig. 2b). The near-infrared Ca II triplet and emission lines due to [Ca II] $\lambda = 7,291$–7,324 Å (strong in core-collapse supernovae and absent in type Ia supernovae) are clearly visible in both objects. The spectrum at +65 days does not show any evidence of the prominent forbidden iron lines which dominate late-time spectra of thermonuclear supernovae. The lack of these features is a strong indication that little $^{56}$Ni was synthesized in the explosion, which does not suggest a thermonuclear origin. The [O I] $\lambda = 6,300$–6,364 Å feature, which is usually prominent in stripped-envelope core-collapse supernovae, is undetected in the spectra of SN 2005cs and SN 2008ha, probably because of the high ejecta density at these phases.

SN 2008ha is the faintest and lowest-luminosity hydrogen-deficient supernova known. Using a distance modulus of 31.55 mag and a reddening of $E(B - V) = 0.076$ mag (see Supplementary Information for details), SN 2008ha has a peak magnitude of $M_R = -14.5 \pm 0.3$. This is five magnitudes fainter than typical type Ia supernovae, and three magnitudes fainter than the low-luminosity thermonuclear explosions. A 'pseudo-bolometric' light curve is shown in Fig. 3a together with light curves of the thermonuclear SN 1991T and other core-collapse supernovae (SN 1998bw[14], SN 2007gr and SN 2005cs). SN 2008ha evolves much more rapidly than these other supernovae and has a maximum luminosity comparable with those of low-luminosity type IIP supernovae[12,15].

The rapidly evolving light curve of SN 2008ha, together with the modest ejecta velocities, implies very low kinetic energy and ejected

[1]Astrophysics Research Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK. [2]INAF Osservatorio Astronomico di Padova, Vicolo dell' Osservatorio 5, I-35122 Padova, Italy. [3]Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, D-85741 Garching bei München, Germany. [4]Begues Observatory, Santpere 6 Casa 22, 08859 Begues, Barcelona, Spain. [5]Spitzer Science Center, California Institute of Technology, 1200 E. California Blvd, Pasadena, California 91125, USA. [6]Calle de la Guardia Civil 22, 46020 Valencia, Spain. [7]Fundación Galileo Galilei-INAF, Telescopio Nazionale Galileo, E-38700 Santa Cruz de la Palma, Tenerife, Spain. [8]Taurus Hill Observatory, Härcämäentie 88, 79480 Kangaslampi, Finland. [9]Tuorla Observatory, Department of Physics and Astronomy, University of Turku, Väisäläntie 20, FI-21500 Piikkiö, Finland. [10]INAF Osservatorio Astronomico di Trieste, Via Tiepolo 11, I-34131 Trieste, Italy. [11]INAF Osservatorio Astronomico di Catania, 78 Via S. Sofia, 95123 Catania, Italy.
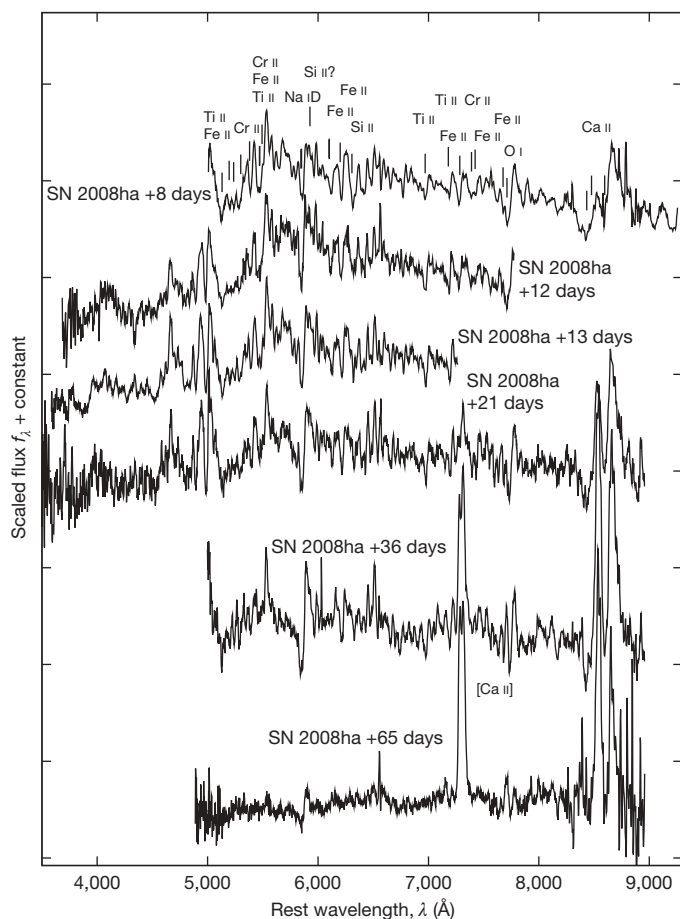
**Figure 1 | Spectral evolution of SN 2008ha.** Spectral evolution of SN 2008ha from +8 to +65 days after R-band maximum (which we estimate to have occurred on Julian day JD = 2454787 ± 2). There is no signature of hydrogen features in the spectra, whereas Fe II, Ti II, Cr II, Na ID, O I and Ca II are clearly detected. The expansion velocity (measured from the minimum of the Fe II lines) is slowly decreasing from 2,300 to 1,500 km s$^{-1}$ in the covered spectral range suggesting a very-slow-velocity evolution. The low velocity of SN 2008ha allows us to identify most lines that are normally blended. Fe II and other metal lines are identified, and those of intermediate-mass elements (Na I, Ca II, O I) are prominent at all epochs (see Supplementary Fig. 3). After +36 days strong lines of [Ca II], typical of core-collapse supernovae, become visible, confirming the fast evolution to the nebular phase of SN 2008ha. The low expansion velocity and the fast spectral evolution suggest an extremely low kinetic energy (1–5 × 10$^{49}$ erg) and ejected mass (0.1–0.5 solar masses) of SN 2008ha. These values are inconsistent with a thermonuclear scenario and quite common in faint core-collapse supernovae.

mass. We roughly estimate these quantities using a toy model based on Arnett's equations[16] (see Supplementary Information), obtaining an ejecta mass of 0.1–0.5 solar masses and a kinetic energy in the range 1–5 × 10$^{49}$ erg. The ejected mass is significantly smaller than the canonical 1.4 solar masses expected for thermonuclear supernovae (ref. 17) (with the caveat that the physical values can be better constrained with more accurate modelling). The kinetic energy is also smaller than suggested by models of the pure deflagration scenario, which were proposed to explain SN 2002cx-like events. We also estimate the mass of $^{56}$Ni produced in the explosion of SN 2008ha to be 0.003–0.005 solar masses (see Supplementary Information). This amount is very close to that observed in sub-luminous type IIP supernovae[12,15], but is two orders of magnitude smaller than in normal type Ia supernovae (0.4–0.8 solar masses of $^{56}$Ni , ref. 18). We cannot exclude that some exotic thermonuclear explosion might be consistent with the observed low energy and fast light curve evolution of SN 2008ha, but the observational comparisons indicate that it
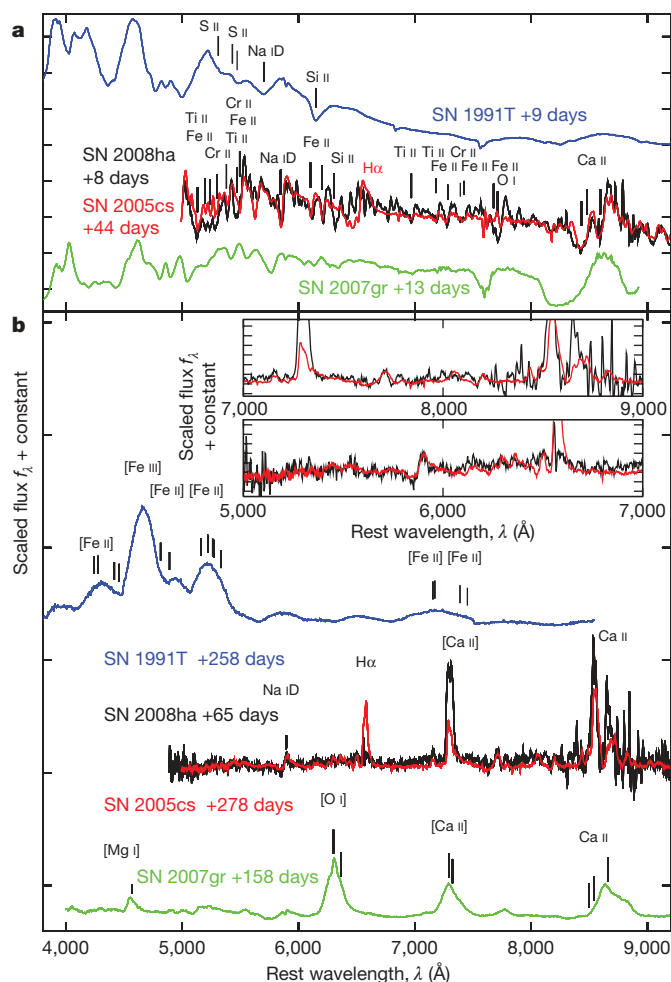


**Figure 2 | Comparison of spectra of SN 2008ha with those of other supernovae. a,** The spectrum of SN 2008ha taken 8 days after maximum is compared with that of the under-luminous type IIP SN 2005cs[15] during the hydrogen recombination, with those of the luminous thermonuclear SN 1991T[29] (this sub-type of type Ia supernovae has been proposed to share some spectral similarities with SN 2002cx[19]) and the type Ic SN 2007gr[13] at comparable phases. The spectra of SN 2008ha and SN 2005cs are very similar, except for the H lines (always prominent in type IIP supernovae) and the O I feature at $\lambda = 7,774$ Å. The photospheric spectra of type Ia and type Ic supernovae both share some similarities with SN 2008ha but only if the spectra are red-shifted (by an *ad hoc* velocity) or smoothed with a gaussian filter. **b,** The latest spectrum of SN 2008ha at 65 days past the R-band maximum has not fully completed the transition to the nebular phase. However the remarkably fast evolution of this event means that it can be compared with the nebular spectra of other supernovae. SN 2008ha is very similar to the nebular spectrum of SN 2005cs[15] except for the distinct lack of hydrogen features (enlarged in the inset plots). The prominent [Fe II] lines typical of thermonuclear supernovae and the [O I] lines of stripped-envelope core-collapse supernovae are not apparent in SN 2008ha.

is more likely that SN 2008ha was produced in the low-energy core-collapse explosion of a hydrogen-deficient massive star.

This discovery has important implications for the origin of some gamma-ray bursts. Several nearby long-duration gamma-ray bursts show evidence of an accompanying bright, highly energetic, envelope-stripped core-collapse supernova in their light curves and spectra. Two long gamma-ray bursts (GRB 060614 and GRB 060505, refs 7–9) were close enough that the presence of an associated supernova could be excluded down to limiting absolute magnitudes of $M_R \approx -12.3$ to $-13.7$. One possible explanation is that they were accompanied by an extremely sub-luminous, hydrogen-poor core-collapse supernova. The discovery of SN 2008ha is the first evidence that such supernovae do exist (see Fig. 3b).
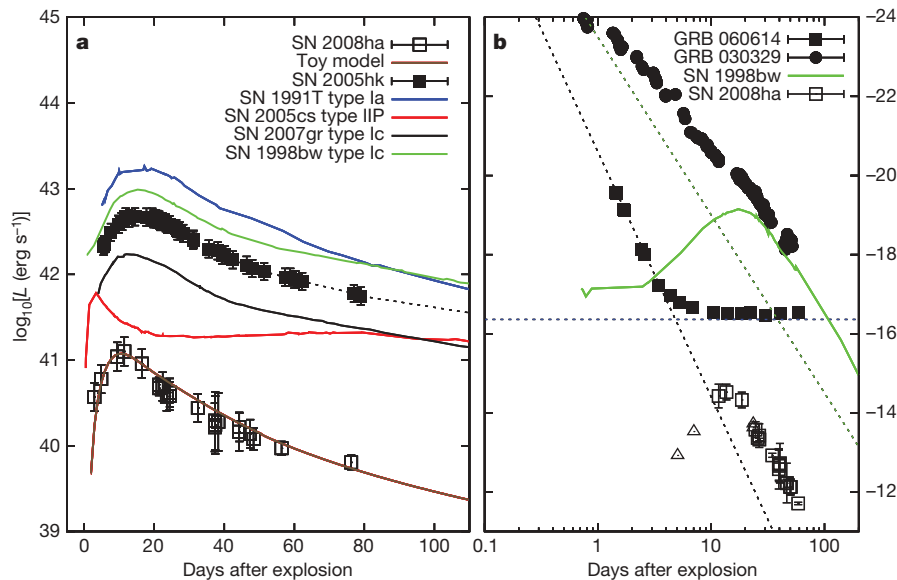
**Figure 3 | Pseudo-bolometric and absolute R-band light curves of SN 2008ha.** **a,** We compare a pseudo-bolometric light curve of SN 2008ha with those of other supernova types (error bars show the standard errors). The pseudo-bolometric light curve of SN 2008ha was computed, using our R-band observations and SN 2005hk as references (after having time-stretched the data of SN 2005hk to fit the time evolution of SN 2008ha). The light curve of SN 2008ha is faster than those of most other supernovae. Considering the low expansion velocity, the light curve of SN 2008ha is inconsistent with a thermonuclear explosion of 1.4 solar masses, typical of type Ia supernovae. The peak luminosity is similar to SN 2005cs, suggesting possibly a stripped-core analogue of this type of explosion[15], or the weak explosion of a very massive star with black-hole formation[1–3] **b,** The R-band light curve of SN 2008ha is shown together with those of SN 1998bw[14] and the afterglows of two long-duration gamma-ray bursts (error bars show the

standard errors). The detection of ref. 10 and magnitudes from the Bright Supernova website (http://www.rochesterastronomy.org/snimages/) are shown as triangles. GRB 060614 showed no evidence for a supernova signature in its R-band afterglow light curve, leading the authors of refs 7, 8 and 9 to suggest that a faint supernova (fainter than −13.7 at peak) would be required to be consistent with the canonical physical production mechanism for long gamma-ray bursts. The horizontal line (blue dotted line) is the host galaxy of GRB 060614. The light curve of SN 2008ha shows, for the first time, that such faint hydrogen poor core-collapse supernovae (even though SN 2008ha was slightly brighter than −13.7 at maximum (being $R = -14.5 \pm 0.3$), do exist. As a comparison, the afterglow of a bright gamma-ray burst (GRB 030329) consistent with the explosion of a SN 1998bw-like event is also shown[30]. In that case the flux excess with respect to the afterglow (green dotted line) was partially due to the SN 1998bw event.
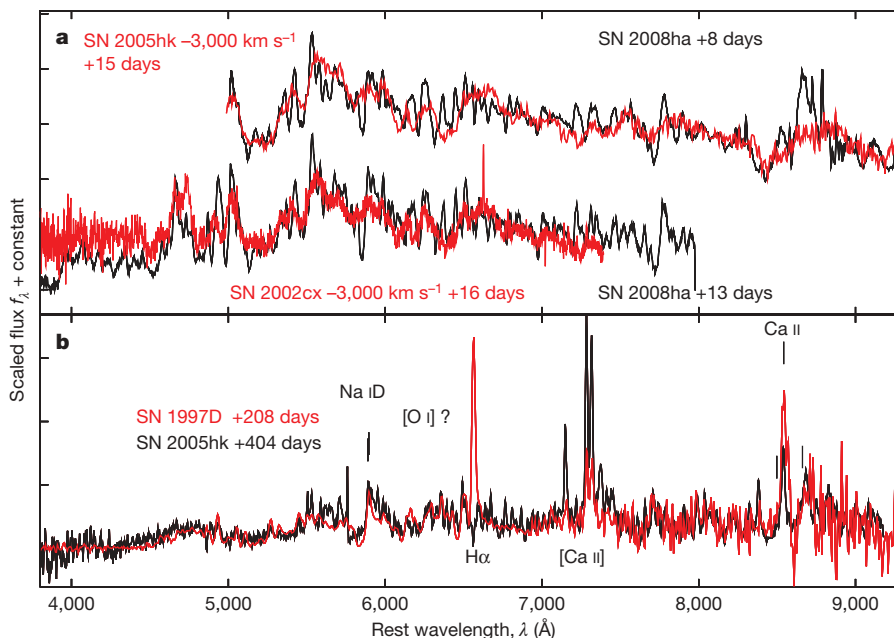


**Figure 4 | Spectra of SN 2008ha and SN 2002cx-like events.** **a,** The spectrum of SN 2008ha is compared with those of SN 2002cx[6,19,20] and SN 2005hk[21,22] at 2–3 weeks after maximum light. The most significant differences between the spectra are due to a different degree of line blending. The spectra of supernovae 2005hk and 2002cx are shifted by −3,000 km s$^{-1}$ to align the absorptions. **b,** The nebular spectrum of SN 2005hk at ∼400 days is compared with that of the faint hydrogen-rich supernova SN 1997D[25]. In analogy to SN 1997D, the nebular spectrum of SN 2005hk is rich in

permitted Fe II lines, clear evidence that the ejecta are still dense and not completely transparent. The [Fe II] and [Fe III] lines typical of thermonuclear supernovae, and the [O I] typical of stripped-envelope core-collapse supernovae, are not detected. In SN 1997D the [O I] feature appeared only about six months later[25]. The spectrum of SN 2005hk was taken on 27 November 2006 at the Very Large Telescope VLT+FORS2 (programme ID 078.D0246) (Stanishev, V. et al., manuscript in preparation).

The observed properties of SN 2008ha are undeniably similar to the group of supernovae similar to SN 2002cx[6,19–22], and it is perhaps the most extreme object of its kind (see Supplementary Information for more details on this supernova group). If SN 2008ha is more plausibly explained by core collapse then, by implication, all SN 2002cx events could possibly be explosions of this nature. So far they have been interpreted as pure thermonuclear deflagrations of 1.4-solar-mass (Chandrasekhar mass) white dwarfs[23], although their observed characteristics deviate significantly from those of type Ia supernovae. Their intrinsic faintness and broad light curves are at odds with the luminosity versus light curve shape relation[24] which characterizes type Ia supernovae and itself is a consequence of comparable ejecta masses (ref. 17). The spectra of SN 2002cx-like supernovae are quite bizarre: before maximum they show similarities to luminous type Ia supernovae[19] (for example, SN 1991T), after maximum they are very similar to those of SN 2008ha (see Fig. 4a) and at late time the spectra resemble those of faint core-collapse supernovae. A comparison of an unpublished late-time spectrum of SN 2005hk[21,22] (the best-studied SN 2002cx-like event), with that of the sub-luminous type IIP SN 1997D[25] is shown in Fig. 4b. The lack of forbidden lines of oxygen and iron in both spectra, and the presence of P-Cygni-type lines of Fe II about 400 days after explosion (in the case of SN 2005hk) is evidence of high density ($\sim 10^8$ cm$^{-3}$, ref. 22). The authors of ref. 22, indeed, attempted to model a spectrum of SN 2005hk at 228 days by combining a photospheric spectrum and a nebular spectrum. However, the model required to reproduce the photospheric phase (a W7 model[26] scaled down to an energy of $3 \times 10^{50}$ erg) was unable to reproduce the high density of the inner ejecta at late phases. High density and low energy are sometimes found in core-collapse supernovae, particularly in faint type IIP supernovae[12,15].

Additional support comes from the detection of He I lines in SN 2007J[27], a supernova showing remarkable similarities with SN 2002cx at an earlier phase[28]. Helium lines have never been detected in thermonuclear supernovae. There is quite significant evidence for SN 2008ha being a core-collapse supernova, and this family of objects could plausibly be of the same nature. Future observations of similar events will help us to understand whether they are a form of thermonuclear explosions, low-luminosity core-collapse supernovae from stripped stars of moderate mass, or the deaths of very massive stars inducing black-hole formation and fall-back.

1. MacFadyen, A. I., Woosley, S. E. & Heger, A. Supernovae, jets, and collapsars. *Astrophys. J.* **550**, 410–425 (2001).
2. Heger, A. *et al.* How massive single stars end their life. *Astrophys. J.* **591**, 288–300 (2003).
3. Fryer, C. L., Young, P. A. & Hungerford, A. L. Explosive nucleosynthesis from gamma-ray burst and hypernova progenitors: direct collapse versus fallback. *Astrophys. J.* **650**, 1028–1047 (2006).
4. Nomoto, K. Evolution of 8–10 solar mass stars toward electron capture supernovae. I. Formation of electron-degenerate O + Ne + Mg cores. *Astrophys. J.* **277**, 791–805 (1984).
5. Kitaura, F. S., Janka, H.-T. h. & Hillebrandt, W. Explosions of O-Ne-Mg cores, the Crab supernova, and subluminous type II-P supernovae. *Astron. Astrophys.* **450**, 345–350 (2006).
6. Li, W. *et al.* SN 2002cx: The most peculiar known type Ia supernova. *Publ. Astron. Soc. Pacif.* **115**, 453–473 (2003).
7. Della Valle, M. *et al.* An enigmatic long-lasting γ-ray burst not accompanied by a bright supernova. *Nature* **444**, 1050–1052 (2006).
8. Gal-Yam, A. *et al.* A novel explosive process is required for the γ-ray burst GRB 060614. *Nature* **444**, 1053–1055 (2006).
9. Fynbo, J. P. U. *et al.* No supernovae associated with two long-duration γ-ray bursts. *Nature* **444**, 1047–1049 (2006).
10. Puckett, T. Moore, C. Newton, J. & Orff. T. Supernova 2008ha in UGC 12682. *Central Bureau Electr. Telegr.* **1567** (2008).
11. Foley, R. J. *et al.* Supernova 2008ha in UGC 12682. *Central Bureau Electr. Telegr.* **1576** (2008).
12. Pastorello, A. *et al.* Low-luminosity type II supernovae: spectroscopic and photometric evolution. *Mon. Not. R. Astron. Soc.* **347**, 74–94 (2004).
13. Valenti, S. *et al.* The carbon-rich type Ic SN 2007gr: the photospheric phase. *Astrophys. J.* **673**, L155–L158 (2008).
14. Patat, F. *et al.* The metamorphosis of SN 1998bw. *Astrophys. J.* **555**, 900–917 (2001).
15. Pastorello, A. *et al.* SN 2005cs in M51. II. Complete evolution in the optical and the near-infrared. *Mon. Not. R. Astron. Soc.* **394**, 2266–2282 (2009).
16. Arnett, W. D. Type I supernovae. I—Analytic solutions for the early part of the light curve. *Astrophys. J.* **253**, 785–797 (1982).
17. Mazzali, P. A., Röpke, F. K., Benetti, S. & Hillebrandt, W. A common explosion mechanism for type Ia supernovae. *Science* **315**, 825–828 (2007).
18. Mazzali, P. A. *et al.* Can differences in the nickel abundance in Chandrasekhar-mass models explain the relation between the brightness and decline rate of normal type IA supernovae? *Astrophys. J.* **547**, 988–994 (2001).
19. Branch, D. *et al.* Reading the spectra of the most peculiar type Ia supernova 2002cx. *Publ. Astron. Soc. Pacif.* **116**, 903–908 (2004).
20. Jha, S. *et al.* Late-time spectroscopy of SN 2002cx: the prototype of a new subclass of type Ia supernovae. *Astron. J.* **132**, 189–196 (2006).
21. Phillips, M. M. *et al.* The peculiar SN 2005hk: do some type Ia supernovae explode as deflagrations? *Publ. Astron. Soc. Pacif.* **119**, 360–387 (2007).
22. Sahu, D. K. *et al.* The evolution of the peculiar type Ia supernova SN 2005hk over 400 days. *Astrophys. J.* **680**, 580–592 (2008).
23. Gamezo, V. N., Khokhlov, A. M. & Oran, E. S. Deflagrations and detonations in thermonuclear supernovae. *Phys. Rev. Lett.* **92**, 211102 (2004).
24. Phillips, M. M. The absolute magnitudes of type IA supernovae. *Astrophys. J.* **413**, L105–L108 (1993).
25. Benetti, S. *et al.* The fading of supernova 1997D. *Mon. Not. R. Astron. Soc.* **322**, 361–368 (2001).
26. Nomoto, K., Thielemann, F. K. & Yokoi, K. Accreting white dwarf models of type I supernovae. III—Carbon deflagration supernovae. *Astrophys. J.* **286**, 644–658 (1984).
27. Filippenko, A. V. *et al.* Supernova 2007J in UGC 1778. *Central Bureau Electr. Telegr.* **926**, 1 (2007).
28. Filippenko, A. V. *et al.* Supernova 2007J in UGC 1778. *Central Bureau Electr. Telegr.* **817**, 1 (2007).
29. Filippenko, A. *et al.* The peculiar type IA SN 1991T—Detonation of a white dwarf? *Astrophys. J.* **384**, L15–L18 (1992).
30. Matheson, T. *et al.* Photometry and spectroscopy of GRB 030329 and its associated supernova 2003dh: the first two months. *Astrophys. J.* **599**, 394–407 (2003).

# LETTERS

# Global circulation as the main source of cloud activity on Titan

Sébastien Rodriguez[1,2], Stéphane Le Mouélic[1], Pascal Rannou[3,4], Gabriel Tobie[1], Kevin H. Baines[5], Jason W. Barnes[6], Caitlin A. Griffith[7], Mathieu Hirtzig[8], Karly M. Pitman[5], Christophe Sotin[1,5], Robert H. Brown[7], Bonnie J. Buratti[5], Roger N. Clark[9] & Phil D. Nicholson[10]

Clouds on Titan result from the condensation of methane and ethane and, as on other planets, are primarily structured by circulation of the atmosphere[1–4]. At present, cloud activity mainly occurs in the southern (summer) hemisphere, arising near the pole[5–12] and at mid-latitudes[7,8,13–15] from cumulus updrafts triggered by surface heating and/or local methane sources, and at the north (winter) pole[16,17], resulting from the subsidence and condensation of ethane-rich air into the colder troposphere. General circulation models[1–3] predict that this distribution should change with the seasons on a 15-year timescale, and that clouds should develop under certain circumstances at temperate latitudes ($\sim$40°) in the winter hemisphere[2]. The models, however, have hitherto been poorly constrained and their long-term predictions have not yet been observationally verified. Here we report that the global spatial cloud coverage on Titan is in general agreement with the models, confirming that cloud activity is mainly controlled by the global circulation. The non-detection of clouds at latitude $\sim$40° N and the persistence of the southern clouds while the southern summer is ending are, however, both contrary to predictions. This suggests that Titan's equator-to-pole thermal contrast is overestimated in the models and that its atmosphere responds to the seasonal forcing with a greater inertia than expected.

The Visual and Infrared Mapping Spectrometer[18] (VIMS) on board NASA's Cassini spacecraft provides a unique opportunity to regularly and accurately chart cloud activity from a close vantage point—hence with high spatial resolution and good spectral coverage. We developed a semi-automated algorithm to isolate clouds from other contributions in VIMS images (Fig. 1) and applied it to 10,000 images of Titan. These images encompass several million spectra, acquired during 39 monthly fly-bys of Titan between July 2004 and December 2007.

The total distribution of cloud events derived from our detections (Fig. 2) and the time variation of their latitudinal distribution (Fig. 3a) indicates that cloud activity is clustered at three distinct latitudes during the 2004–2007 period: the south polar region (poleward of 60° S), the north polar region (poleward of 50° N) and a narrow belt centred at $\sim$40° S. Individual detection maps are provided for each fly-by in Supplementary Figs 1–4.

Our study clearly shows the stability of the north polar cloud, which is systematically detected over the 2004–2007 period. We observe this extensive meteorological system poleward of 50–60° N. All of these clouds spectrally differ from the southern clouds, which are presumably formed by wet convection and are made of large (tens of micrometres in size), liquid/solid methane droplets[2,16]. They produce much less

signal at 5 μm than any other cloudy features we detect elsewhere on Titan, indicating less backscattering at that wavelength. Given that complex indices of refraction of methane and ethane are not that different at 5 μm, the difference in backscattering comes essentially from the particle size. Relatively less backscattering at 5 μm is consistent with north polar clouds composed of smaller (micrometre-sized) particles more probably made of solid ethane[2,16,17]. We also detect small, elongated clouds at $\sim$60–70° N in March and April 2007. Surrounded by the large north polar ethane cloud, these clouds are thought to be convective methane clouds connected to the underlying lakes[19]. Their higher brightness at 5 μm confirms that they are similar to the methane clouds found in the southern hemisphere.

A few tropical clouds, thought to be rare during Titan's summer, are detected close to the equator ($\sim$15° S) on 12 December 2006. Their areas never exceed 10,000 km². These clouds therefore could not have been detected from ground-based observations. More details about tropical clouds are given in ref. 20. We also observe more than 100 isolated and transient temperate clouds near 40° S (Figs 2 and 3a). Most of them are elongated in the east–west direction, as was previously reported[7,8,13–15], possibly owing to orographic waves over zonally oriented topography and/or shearing and stretching by strong zonal winds of tens of metres per second[7]. This type of cloud appeared during two periods, in 2004 and then regularly (in two-thirds of the fly-bys) between July 2006 and October 2007. Between December 2004 and August 2006, temperate clouds are observed very rarely (only in October 2005 (ref. 10) and January 2006 (this study)). This could be attributed to less frequent fly-bys of Titan by Cassini and/or a momentary decline in cloud activity.

We compared our latitudinal and time distribution of clouds (Fig. 3a) with predictions of the IPSL-TGCM[2], which so far is the only general circulation model for Titan to include a microphysical cloud scheme and, thus, predict the cloud cover (Fig. 3b). Except for the lack of winter mid-latitude clouds (40° N), we find that the main spatial characteristics of our cloud distribution are well reproduced by the IPSL-TGCM. Clouds appear in the model near altitude 12 km at latitude $\sim$40° in the summer hemisphere (the southern hemisphere until 2009), and are associated with the ascent of the convergence zone of a Hadley-type cell[1–3]. Clouds are also predicted very near the summer pole, where methane, driven from the warmer region below, condenses, generating convective structures[2,21–23]. In the winter polar region, the cloud formation is related to the downwelling stratospheric circulation, which drives ethane- and aerosol-enriched stratospheric air into the cold tropopause of the polar night (above 40 km). The

[1]Laboratoire de Planétologie et Géodynamique, CNRS, UMR 6112, Université de Nantes, 2 rue de la Houssinière, 44000 Nantes, France. [2]Laboratoire AIM, Université Paris 7, CNRS UMR-7158, CEA-Saclay/DSM/IRFU/SAp, 91191 Gif/Yvette, France. [3]Groupe de Spectrométrie Moléculaire et Atmosphérique, CNRS UMR-6089, Université de Reims Champagne-Ardenne, 51687 Reims, France. [4]LATMOS, CNRS UMR-7620, Université Versailles-St-Quentin, 91371 Verrières-le-Buisson, France. [5]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109-8099, USA. [6]NASA Ames Research Center M/S 244-30, Moffett Field, California 94035, USA. [7]Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721-0092, USA. [8]AOSS, PSL, University of Michigan, Ann Arbor, Michigan 48109, USA. [9]USGS, Denver Federal Center, Denver, Colorado 80225, USA. [10]Cornell University, Astronomy Department, Ithaca, New York 14853, USA.

Examples of Titan's clouds imaged by VIMS in the 2004–2007 period
(RGB composites: red, 5 μm; green, 1.6 μm; blue, 1.27 μm)



South polar cloud    North polar cloud    Various southern temperate clouds
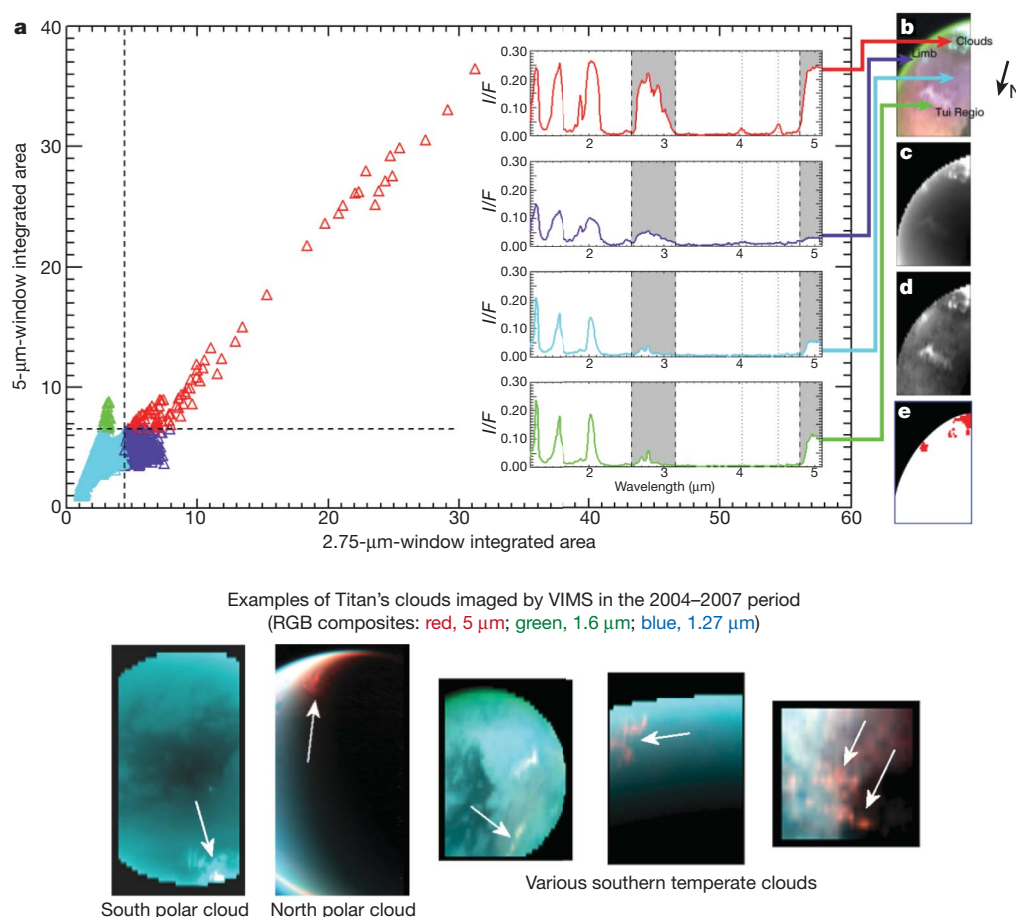
**Figure 1 | Method of spectral detection of Titan's clouds illustrated on a representative VIMS data cube.** The VIMS on board Cassini acquires a 352-channel spectrum from 0.3 to 5.1 μm for each pixel of an image[18]. **a**, Scatter plot of the 2.75-μm-window integrated area versus the 5-μm-window integrated area of the VIMS colour image shown in **b**, with red indicating a wavelength of 2.03 μm; green, 2.78 μm; and blue, 5 μm. The window integrated areas correspond to the integral of $I/F$ within the spectral range shown in grey in the spectra ($I$ is the observed specific intensity and $\pi F$ is the incident solar intensity). **c**, **d**, Images of the 2.75-μm-window (**c**) and 5-μm-window (**d**) integrated areas, coded in greyscale (high values appear bright). Characteristic spectra are inset in **a**, showing clouds (red), limb (violet), typical surface (cyan) and a high 5-μm signal surface feature (Tui Regio[27], green). 'Surface' windows correspond to peaks at 1.27, 1.59, 2.03, 2.75 and

5 μm. Because clouds are efficient reflectors and reduce the path lengths of solar photons, their spectra present a brightening of all surface windows relative to other spectra. We found that the most robust spectral criterion to separate cloud pixels from other contributions (surface and limb) is the simultaneous increased integrated areas of the 2.75- and 5-μm windows. Conservative, $2\sigma$ thresholds for the integrated areas of these two windows are automatically calculated to isolate pixels corresponding to clouds (red triangles in **a**). We deliberately choose a conservative threshold to avoid false positives. This can lead to the rare non-detection of optically thin or low-altitude clouds, of clouds much smaller than a VIMS pixel or of clouds that are too close to the limb. **e**, The resulting cloud-pixel detection (in red), which we reproject on a global map (Fig. 2).

observed stability of the north polar clouds is interpreted, using the IPSL-TGCM, as the result of a constant incoming flux of ethane and aerosols from the stratosphere[24], producing a mist of micrometre-sized droplets of ethane and other products that slowly settles.

However, present observations do not confirm the clouds at ~40° N predicted by the IPSL-TGCM. In the model, these clouds should result from horizontal diffusive transport by inertial instabilities of air, partially humid (relative humidity, 50%) in tropical regions, towards the colder north pole. At an altitude of 12 km, where these clouds are formed, the model predicts an equator-to-pole temperature contrast of $T_{80° N} - T_{0°} = -4$ K. Such a contrast causes the air to become saturated and to produce clouds around 40° N. The lack of such clouds in observations could be explained by an actual equator-to-pole tem-perature contrast of about $-1.5$ K instead of the $-4$ K predicted by the IPSL-TGCM. Such a small thermal contrast would allow air parcels with relative humidities of 50% in tropical regions to move towards the pole without condensing. Conversely, it could also enable the north polar region, where lakes are observed and which is saturated in methane, to wet the tropical regions up to 50% humidity. If we

consider the conditions at the surface, computations, including phase equilibrium with a $N_2$–$CH_4$ mixture, show that with an equator-to-pole temperature contrast near the ground of $-4.2$ K (instead of $-6.5$ K in the IPSL-TGCM), an air parcel at methane saturation near the pole (fed by lakes) would be at 50% humidity if transported at the tropics. Only 80% humidity would be needed at the north pole if the temperature contrast at the surface were to drop to $-3$ K, which is actually observed[25].

In addition, the timing of the summer-hemisphere clouds as con-strained by our observations is poorly reproduced by the IPSL-TGCM. Figure 3b shows that the southern cloud activity should gradually decrease as the equinox approaches, as a consequence of a progressive change in the south polar circulation pattern. This fore-casted decline of southern meteorological activity is not supported by our data. According to the IPSL-TGCM, the south polar clouds should have disappeared in mid-2005 and the mid-latitude clouds should have gradually disappeared since 2005, whereas in our observations the southern clouds are still present even late in 2007 and are particu-larly active at 40° S until mid-2007. The significant latency in the
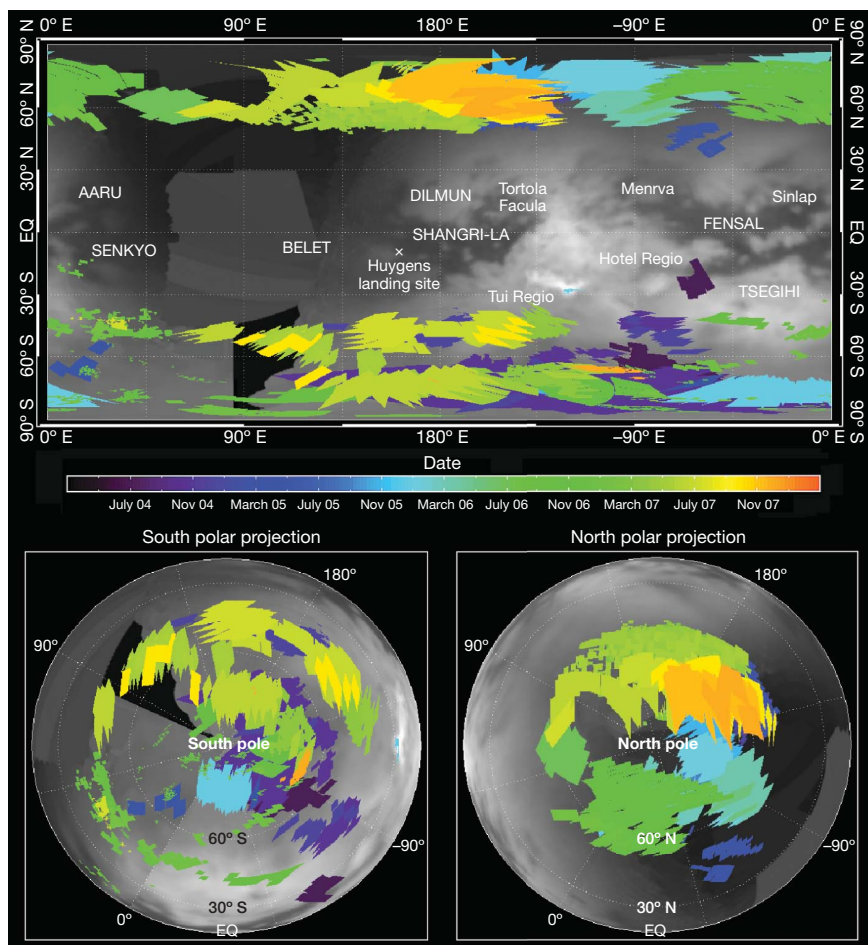
**Figure 2 | Maps of Titan's clouds derived from VIMS observations from July 2004 to December 2007.** Our detections are presented in cylindrical (top) and polar orthographic (bottom) projections. The colours of the clouds correspond to the date of each cloud observation. A VIMS greyscale mosaic of Titan's surface (adapted from RGB colour-composite global mosaics in ref. 28) is used as background. Clouds are found to be distributed in three clustered regions: the two poles and the southern temperate latitudes. Only very few occurrences of clouds are found in equatorial regions. One cloud event is found on December 2005 just above particularly interesting terrain thought to be of cryovolcanic origin (Tui Regio[27]), and may indicate recent cryovolcanic activity. EQ, equator.
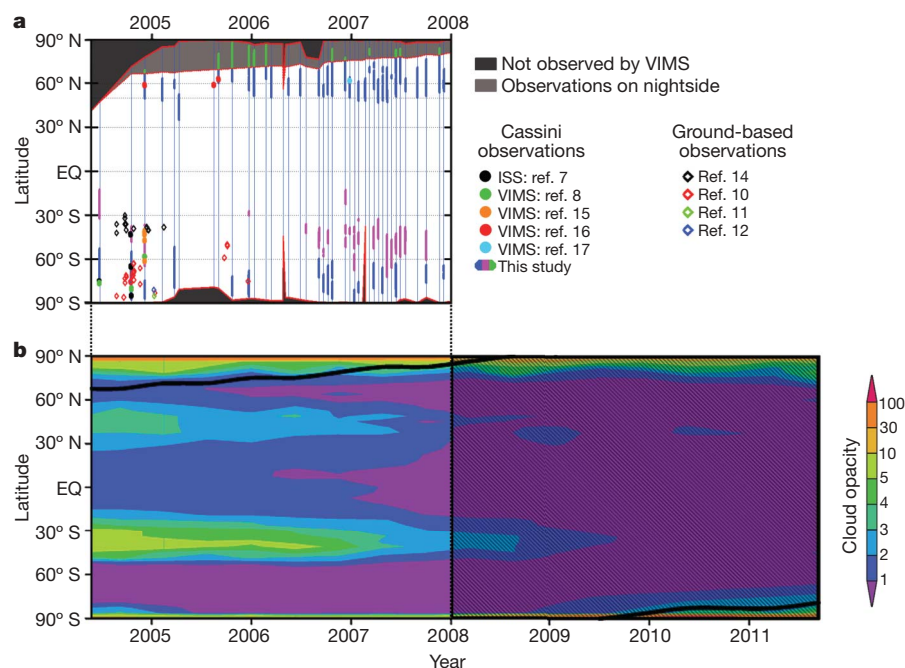


**Figure 3 | Titan's latitudinal cloud coverage over time compared with global circulation model[3] predictions. a**, Latitudinal distribution of clouds detected using VIMS, plotted as a function of time between July 2004 and December 2007. The thin blue vertical lines mark the time of the VIMS observations. The latitude extent of the clouds we detect is enhanced with thicker vertical lines, in blue when on dayside and in green when in polar night. Isolated temperate clouds are coloured purple. The previous Cassini and ground-based observations reported in the literature are superimposed over our latitudinal distribution using coloured dots and diamonds, respectively. Our detections are in very good agreement with the previous observations. ISS, Imaging Science Subsystem (on board Cassini). **b**, Titan's integrated cloud opacity above 10 km, summed each year, predicted by the atmospheric global circulation model of ref. 2 (IPSL-TGCM) between 2004 and 2011. The thick black lines show the edge of the polar night. The spatial distribution of clouds forecasted by the IPSL-TGCM, confining clouds at the two poles and around 40° S, is in very good agreement with our observations (see **a** and Fig. 2). By contrast, the observed cloud timing is poorly reproduced by the IPSL-TGCM. In the time interval monitored using VIMS for this work, the IPSL-TGCM predict that the south polar cloud should vanish before the equinox for more than one year, and that the cloud belt at 40° S should reach a maximum of intensity between 2004 and 2007 and then should gradually vanish with the incoming circulation turnover. This seems to be observed by VIMS with a significant delay (see text).
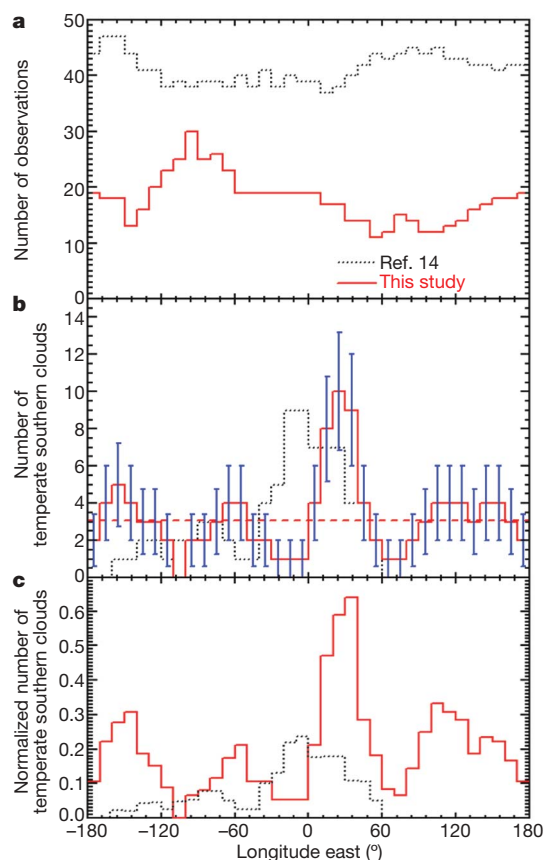
**Figure 4 | The southern temperate cloud distribution as a function of longitude. a,** The total number of observations in each 10° longitude bin is shown by the solid red line for our study and by the black dotted line for ref. 14. **b,** The number of clouds observed by VIMS between July 2004 and December 2007 (our study, solid red line) and during ground-based observations between December 2003 and February 2005 (ref. 14, black dotted line) in each 10° bin of planetocentric longitude summed between 60° S and 0°. Blue bars indicate the Poisson standard deviation for each VIMS cloud count. The statistics indicate that the overall shape of the longitudinal distribution is significant. **c,** Comparison of normalized number of clouds (number of clouds divided by the number of observations) from ref. 14 and from this study. Our distribution shows two minima, at the sub-Saturn point (0° E)—where ref. 14 saw a maximum—and at the anti-Saturn point (180° E). Two other minima are also present, in the neighbourhoods of 70° E and −110° E. Owing to the limitation of Cassini's Saturn tour, the detection of clouds was heavily precluded here by particularly low spatial resolution (Supplementary Fig. 5a) and very unfavourable observation conditions (resulting from high air mass; Supplementary Fig. 5b), so these two minima cannot be interpreted with confidence.

predicted disappearance of summer clouds suggests that Titan's atmosphere responds to seasonal forcing with a greater inertia than expected. Since August 2007, however, the occurrence of south polar clouds seems to be less frequent in our data and the mid-latitude clouds seem to be scarcer. These very subtle declining trends may indicate that we are witnessing the approaching seasonal circulation turnover as we approach the equinox, but with a timing pattern different from that forecasted by the IPSL-TGCM.

Besides, Fig. 4 shows that, between July 2004 and December 2007, the mid-latitude clouds are not uniformly distributed in longitude, as already noticed during previous ground-based observations[14] (December 2003–February 2005). The propensity of the clouds to form around longitude 0° found in 2003–2005 was attributed to localized geological forcings from the surface that are possibly related to an active cryovolcanic province[14]. However, three years later, the distribution we

measure differs markedly, showing more structure (Fig. 4c). In contradiction to ref. 14, we observe mid-latitude clouds at almost all longitudes, with an excess at longitudes (from 60° E to 180° E, which corresponds to the leading hemisphere of Titan) where ref. 14 detected none. The strong and secondary peaks in the number of clouds reported by ref. 14 have drifted eastward by 30° at an estimated rate of ~10° per terrestrial year. In addition, we found two troughs at longitudes facing towards (0°) and away from (180°) Saturn. Although the strong link between cloud number and latitude indicates that global circulation has a major role in cloud formation[1–3], the wavy pattern of our cloud distribution suggests a secondary forcing mechanism. The longitudinal shift in cloud distribution by 30° between 2003–2004 (ref. 14) and 2005–2007 (this study), as well as the loose correlation of clouds with surface location, excludes surface geological activity as the primary triggering mechanism.

Both the drift in longitude and the discovery of two diametrically opposite minima favour processes taking place in Titan's atmosphere, which we attribute to external forcing by Saturn's tides. Saturn's tides are predicted to generate tidal winds in Titan's dense atmosphere that are particularly significant in the troposphere[26] at altitudes where temperate clouds are found to develop[2,3,13–15]. These winds manifest themselves as east-moving planetary-scale waves of degree two and change west–east direction periodically during each tidally locked orbit of Titan[26]. Consequently, tidally induced winds periodically modify the convergence of air masses, mostly at two preferential longitudes 180° apart, potentially resulting in perturbations to cloud formations[26].

The extension of the Cassini mission possibly up to the summer solstice in 2017, and the continuation of ground-based observations, will provide further observational constraints on the general circulation models. The refined models should provide more accurate information about the global atmospheric circulation, which is crucial for understanding the carbon cycle on Titan.

1. Tokano, T. Three-dimensional modeling of the tropospheric methane cycle on Titan. *Icarus* **153,** 130–147 (2001).
2. Rannou, P., Montmessin, F., Hourdin, F. & Lebonnois, S. The latitudinal distribution of clouds on Titan. *Science* **311,** 201–205 (2006).
3. Mitchell, J. L., Pierrehumbert, R. T., Frierson, D. M. W. & Caballero, R. The dynamics behind Titan's methane clouds. *Proc. Natl Acad. Sci. USA* **103,** 18421–18426 (2006).
4. Atreya, S. K. *et al.* Titan's methane cycle. *Planet. Space Sci.* **54,** 1177–1187 (2006).
5. Brown, M. E., Bouchez, A. H. & Griffith, C. A. Direct detection of variable tropospheric clouds near Titan's south pole. *Nature* **420,** 795–797 (2002).
6. Bouchez, A. H. & Brown, M. E. Statistics of Titan's south polar tropospheric clouds. *Astrophys. J.* **618,** L53–L56 (2005).
7. Porco, C. C. *et al.* Imaging of Titan from the Cassini spacecraft. *Nature* **434,** 159–168 (2005).
8. Baines, K. H. *et al.* The atmospheres of Saturn and Titan in the near-infrared: first results of Cassini/VIMS. *Earth Moon Planet.* **96,** 119–147 (2005).
9. Schaller, E. L., Brown, M. E., Roe, H. G. & Bouchez, A. H. A large cloud outburst at Titan's south pole. *Icarus* **182,** 224–229 (2006).
10. Schaller, E. L., Brown, M. E., Roe, H. G., Bouchez, A. H. & Trujillo, C. A. Dissipation of Titan's south polar clouds. *Icarus* **184,** 517–523 (2006).
11. de Pater, I. *et al.* Titan imagery with Keck adaptive optics during and after probe entry. *J. Geophys. Res.* **111,** doi:10.1029/2005JE002620 (2006).
12. Hirtzig, M. *et al.* Monitoring atmospheric phenomena on Titan. *Astron. Astrophys.* **456,** 761–774 (2006).
13. Roe, H. G., Bouchez, A. H., Trujillo, C. A., Schaller, E. L. & Brown, M. E. Discovery of temperate latitude clouds on Titan. *Astrophys. J.* **618,** L49–L52 (2005).
14. Roe, H. G., Brown, M. E., Schaller, E. L., Bouchez, A. H. & Trujillo, C. A. Geographic control of Titan's mid-latitude clouds. *Science* **310,** 477–479 (2005).
15. Griffith, C. A. *et al.* The evolution of Titan's mid-latitude clouds. *Science* **310,** 474–477 (2005).
16. Griffith, C. A. *et al.* Evidence for a polar ethane cloud on Titan. *Science* **313,** 1620–1622 (2006).
17. Le Mouélic, S. *et al.* Imaging of the North polar cloud on Titan by the VIMS Imaging Spectrometer onboard Cassini. *Lunar Planet. Sci. Conf.* **39,** 1649–1650 (2008).
18. Brown, R. H. *et al.* The Cassini Visual and Infrared Mapping Spectrometer investigation. *Space Sci. Rev.* **115,** 111–168 (2004).
19. Brown, M. E. *et al.* Discovery of lake-effect clouds on Titan. *Geophys. Res. Lett.* **36,** doi:10.1029/2008GL035964 (2009).

20. Griffith, C. A. *et al.* Characterization of clouds in Titan's tropical atmosphere. *Science.* (submitted).
21. Hueso, R. & Sanchez-Lavega, A. Methane storms on Saturn's moon Titan. *Nature* **442,** 428–431 (2006).
22. Barth, E. L. & Rafkin, S. C. R. TRAMS: a new dynamic cloud model for Titan's methane clouds. *Geophys. Res. Lett.* **34,** doi:10.1029/2006GL028652 (2007).
23. Turtle, E. P. *et al.* Cassini imaging of Titan's high-latitude lakes, clouds, and south-polar surface changes. *Geophys. Res. Lett.* **36,** doi:10.1029/2008GL036186 (2009).
24. McKay, C. P. *et al.* Physical properties of the organic aerosols and clouds of Titan. *Planet. Space Sci.* **49,** 79–99 (2001).
25. Jennings, D. E. *et al.* Titan's surface brightness temperatures. *Astrophys. J.* **691,** L103–L105 (2009).
26. Tokano, T. & Neubauer, F. M. Tidal winds on Titan caused by Saturn. *Icarus* **158,** 499–515 (2002).
27. Barnes, J. W. *et al.* A 5-micron-bright spot on Titan: evidence for surface diversity. *Science* **310,** 92–95 (2005).
28. Barnes, J. W. *et al.* Global-scale surface spectral variations on Titan seen from Cassini/VIMS. *Icarus* **186,** 242–258 (2007).

# LETTERS

# Entangled mechanical oscillators

J. D. Jost[1], J. P. Home[1], J. M. Amini[1], D. Hanneke[1], R. Ozeri[2], C. Langer[3], J. J. Bollinger[1], D. Leibfried[1] & D. J. Wineland[1]

Hallmarks of quantum mechanics include superposition and entanglement. In the context of large complex systems, these features should lead to situations as envisaged in the 'Schrödinger's cat'[1] thought experiment (where the cat exists in a superposition of alive and dead states entangled with a radioactive nucleus). Such situations are not observed in nature. This may be simply due to our inability to sufficiently isolate the system of interest from the surrounding environment[2,3]—a technical limitation. Another possibility is some as-yet-undiscovered mechanism that prevents the formation of macroscopic entangled states[4]. Such a limitation might depend on the number of elementary constituents in the system[5] or on the types of degrees of freedom that are entangled. Tests of the latter possibility have been made with photons, atoms and condensed matter devices[6,7]. One system ubiquitous to nature where entanglement has not been previously demonstrated consists of distinct mechanical oscillators. Here we demonstrate deterministic entanglement of separated mechanical oscillators, consisting of the vibrational states of two pairs of atomic ions held in different locations. We also demonstrate entanglement of the internal states of an atomic ion with a distant mechanical oscillator. These results show quantum entanglement in a degree of freedom that pervades the classical world. Such experiments may lead to the generation of entangled states of larger-scale mechanical oscillators[8–10], and offer possibilities for testing non-locality with mesoscopic systems[11]. In addition, the control developed here is an important ingredient for scaling-up quantum information processing with trapped atomic ions[12–14].

Mechanical oscillators pervade nature; examples include the vibrations of violin strings, the oscillations of quartz crystals used in clocks, and the vibrations of atoms in a molecule. Independent of the size of the system, each mode of vibration can be described by the same equations that describe the oscillations of a mass attached to a fixed object by a spring. For very low energy oscillations, quantum mechanics is needed for a correct description: the energy is quantized, and the motion can be described generally by superpositions of wavefunctions corresponding to each quantum level. Coherent states behave very much like classical oscillators, while other states have properties with distinctly non-classical features[15]. Quantum mechanics also permits superposition states of multiple systems called entangled states, where the measured properties of the systems are correlated in ways that defy our everyday experience[6,7,16–18]. When extended to macroscopic scales, situations akin to Schrödinger's cat should appear. Our inability to produce such macroscopic entanglement may be just a question of technical difficulty. However, there might be a more fundamental cause, such as the inability to entangle certain types of degrees of freedom.

To explore the latter territory in a new regime, we demonstrate entanglement of two separated mechanical oscillators. Here each oscillator comprises a pair of ions—one $^9Be^+$ and one $^{24}Mg^+$—confined in a potential well. In the context of the experiment described below, each pair behaves like two masses connected by a spring of length $\sim4\,\mu m$,

undergoing vibrational motion (Fig. 1). The two pairs are separated by $\sim0.24\,mm$ such that the coupling between them can be neglected. To create the entangled state of the oscillators, we start with all four ions in one location and entangle the internal states of the two $^9Be^+$ ions[19]. We then separate the four ions into two pairs, each containing one of the entangled $^9Be^+$ ions. Finally, we transfer the entanglement from the $^9Be^+$ ions' internal states to the motion of the separated ion pairs, creating the desired motional entanglement.

Initially, all the ions are held in a single potential well of a multi-zone linear Paul trap[20,21]. The potential well is configured to locate the ions along a line corresponding to the axis of weakest confinement, which we call the axial direction. We will be concerned only with motional modes along this axis. While applying continuous laser cooling, we initialize the ions in a particular order, $^9Be^+$–$^{24}Mg^+$–$^{24}Mg^+$–$^9Be^+$, by first increasing the axial confinement until no linear arrangement is stable. The axial potential is independent of ion mass while the radial potential strength scales inversely with the mass[12], thus there exist axial potentials where the heavier $^{24}Mg^+$ ions are displaced from the axis and must reside between the $^9Be^+$ ions. We then relax the axial confinement giving the desired order (T. Rosenband, personal communication).

Lasers provide control of the ions' motion and internal states through laser cooling and stimulated-Raman carrier or sideband transitions[7,12,22]. Using Doppler cooling on the $^9Be^+$ and $^{24}Mg^+$ ions, followed by sideband cooling on the $^9Be^+$ ions, we prepare the motion of each of the four axial normal modes (see Methods) to an average motional occupation of $\langle n \rangle \lesssim 0.17$. By applying a magnetic field of 0.012 T, we spectrally isolate two internal (hyperfine) states in each $^9Be^+$ ion, which we call 'spin' states, and label $|\uparrow\rangle \equiv |F = 2, m_F = 2\rangle$ and $|\downarrow\rangle \equiv |F = 2, m_F = 1\rangle$, where $F$ is the ion's total angular momentum, and $m_F$ its projection along the magnetic field direction. These states are split by 102 MHz. Using a geometric phase gate[19] and spin rotations, we create the decoherence-free entangled state[23]

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}[|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle] \tag{1}$$

of the two $^9Be^+$ ions. This state is resistant to decoherence from spatially uniform magnetic field noise.



**Figure 1 | Mechanical oscillators.** Simplified depiction of the two mechanical oscillators indicating motion in the stretch mode of each $^9Be^+$–$^{24}Mg^+$ ion pair, held in separate locations (not to scale). The pairs—spaced by $\sim0.24\,mm$—each behave as two masses spaced by $\sim4\,\mu m$, connected by a spring. Blue, $^{24}Mg^+$ ions; red, $^9Be^+$ ions.

[1]Time and Frequency Division, National Institute of Standards and Technology, Boulder, Colorado 80305, USA. [2]Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel. [3]Lockheed Martin, Denver, Colorado 80127, USA.

Time-varying axial potentials move and separate[20,21] the four ions into two $^9\text{Be}^+$–$^{24}\text{Mg}^+$ pairs in different wells, which are spaced by ~0.24 mm (see Fig. 2). Each pair of ions has two axial normal modes: the 'stretch' mode (frequency ~4.9 MHz) in which the two ions oscillate out-of-phase, and the 'common' mode (~2.3 MHz) where they oscillate in-phase. The experiment involves the ground $|n=0\rangle_j$ and first excited $|n=1\rangle_j$ states of the stretch modes, where $j \in \{A, B\}$ refers to the well. In general, the separation process excites the motional modes into unknown states. The wavefunction of the $^9\text{Be}^+$ spin states after separation is

$$\frac{1}{\sqrt{2}}\left[|\uparrow\rangle_A|\downarrow\rangle_B + e^{i\xi(t)}|\downarrow\rangle_A|\uparrow\rangle_B\right] \qquad (2)$$

where $\xi(t)$ is a phase that accumulates through the course of the experiment due to a small difference in magnetic field between wells A and B.

To create the motional entangled state, we first prepare the stretch modes close to $|0\rangle_A|0\rangle_B$. For this, Doppler and sideband laser cooling on the $^{24}\text{Mg}^+$ ions in both wells sympathetically cools[24] the $^9\text{Be}^+$ ions and prepares the stretch modes to mean occupation numbers of $\langle n_A\rangle = 0.06(2)$ and $\langle n_B\rangle = 0.02(2)$, where the uncertainties are standard errors (s.e.m.). We also cool the common mode in each well to $\langle n\rangle \le 0.13$. The cooling does not affect the spin states of the $^9\text{Be}^+$ ions[24], thereby approximating the state:

$$\frac{1}{\sqrt{2}}\left[|\uparrow\rangle_A|\downarrow\rangle_B + e^{i\xi(t)}|\downarrow\rangle_A|\uparrow\rangle_B\right]|0\rangle_A|0\rangle_B \qquad (3)$$

We transfer the entanglement from the spin to the motion with a sequence of laser pulses on the $^9\text{Be}^+$ ions. Carrier transitions (labelled with superscript c, duration ~4 μs) only affect the spin states, and sideband transitions (superscript m, referred to as spin ↔ motion transfer pulses, duration ~13 μs) couple the spin and motion. These can be described as generalized rotations:

$$R_j^{c,m}(\theta,\phi) = \begin{pmatrix} \cos\frac{\theta}{2} & -ie^{-i\phi}\sin\frac{\theta}{2} \\ -ie^{i\phi}\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}$$

where $j \in \{A, B\}$. Carrier transitions correspond to rotations in the basis:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = |\uparrow\rangle, \begin{pmatrix} 0 \\ 1 \end{pmatrix} = |\downarrow\rangle$$
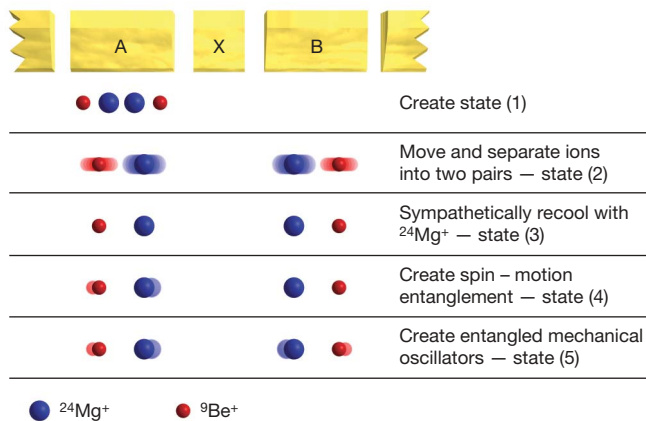


**Figure 2 | Creation of entangled mechanical oscillators.** Schematic showing the ions' positions with respect to the ion trap electrodes (A, X and B) and the quantum states at key points in the experiments (not to scale). After entangling the $^9\text{Be}^+$ ions' spins in a single well, the ions are separated into two pairs by electrode X and distributed to different wells. Laser cooling of $^{24}\text{Mg}^+$ removes motional excitation incurred during separation. A $^9\text{Be}^+$ laser pulse in well A entangles the motion in well A with the $^9\text{Be}^+$ spin in well B. A subsequent pulse in well B entangles the two mechanical oscillators.

and sideband transitions correspond to rotations in the basis:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = |\uparrow\rangle|1\rangle, \begin{pmatrix} 0 \\ 1 \end{pmatrix} = |\downarrow\rangle|0\rangle$$

The rotation angle $\theta$ is proportional to the intensity and duration of the pulses, and the phase $\phi$ is determined by the phase difference between the two optical Raman fields[12,22] at the position of the ion. We individually address the $^9\text{Be}^+$ ions in each well using acousto-optic modulators to shift the positions of the laser beams.

Applying $R_A^m(\pi,0)$ to state (3) entangles the $^9\text{Be}^+$–$^{24}\text{Mg}^+$ motion in well A with the $^9\text{Be}^+$ spin in well B, creating the state:

$$\frac{1}{\sqrt{2}}|\uparrow\rangle_A\left[|\downarrow\rangle_B|0\rangle_A - ie^{i\xi(t)}|\uparrow\rangle_B|1\rangle_A\right]|0\rangle_B \qquad (4)$$

After this spin → motion transfer, the spin in well B is sensitive to decoherence from fluctuating magnetic fields. To minimize this effect, we apply a spin-echo pulse[25], $R_B^c(\pi,0)$, $T \approx 40$ μs after the previous pulse. After a second delay $T$, we apply a second spin → motion transfer pulse $R_B^m(\pi,0)$ in well B, producing the state:

$$\frac{1}{\sqrt{2}}|\uparrow\rangle_A|\uparrow\rangle_B\left[|0\rangle_A|0\rangle_B - e^{i\xi(t)}|1\rangle_A|1\rangle_B\right] \qquad (5)$$

This state is an entangled superposition of both stretch modes in the ground and first excited states. The entanglement now resides only in the mechanical oscillator states of both wells. We leave the system in this state for ~50 μs before beginning our analysis.

We are not able to directly measure the entangled motional state. The analysis proceeds by basically reversing the steps used to create state (5) and characterizing the resulting spin state. We transfer the motional state back into the spins using the pulse sequence: $R_B^m(\pi,0)$, $T$, $R_B^c(\pi,0)$, $T$, $R_A^m(\pi,\phi_A)$. We then recombine all the ions into a single potential well, to ideally reproduce the state $|\Psi^+\rangle$, having chosen $\phi_A$ to compensate for the phase $\xi(t)$.

Imperfect creation of the state (5) could leave entanglement in the spin states, which could mimic motional entanglement in the analysis. To prevent this, we transfer residual populations $\varepsilon_{A,B}$ of states $|\downarrow\rangle_{A,B}$ into auxiliary internal (hyperfine) states before performing the motion → spin transfers (Methods). Moreover, since all experiments are used in the analysis, this transfer process does not constitute postselection and cannot enhance the deduced entanglement.

Our detection relies on analysing the state $|\Psi_f\rangle = \frac{1}{\sqrt{2}}[|\uparrow\uparrow\rangle + i|\downarrow\downarrow\rangle]$, which we create by applying a common rotation $R^c\left(\frac{\pi}{2}, -\frac{3\pi}{4}\right)$ to both spins. We verify the entanglement created in state (5) by measuring the off-diagonal element $|\rho_{\downarrow\downarrow,\uparrow\uparrow}| = |\langle\downarrow\downarrow|\rho_f|\uparrow\uparrow\rangle|$ of the density matrix $\rho_f$ corresponding to our approximation to the state $|\Psi_f\rangle$. We determine $|\rho_{\downarrow\downarrow,\uparrow\uparrow}|$ by applying a final analysis pulse, $R^c\left(\frac{\pi}{2},\phi_p\right)$, to both $^9\text{Be}^+$ ions with a phase $\phi_p$ and measuring the parity[26], $P_{\downarrow\downarrow} + P_{\uparrow\uparrow} - (P_{\downarrow\uparrow} + P_{\uparrow\downarrow})$, for different values of $\phi_p$, where $P_{\downarrow\downarrow}$, $P_{\uparrow\uparrow}$, $P_{\downarrow\uparrow}$ and $P_{\uparrow\downarrow}$ are the populations of the spin states $|\downarrow\downarrow\rangle$, $|\uparrow\uparrow\rangle$, $|\downarrow\uparrow\rangle$ and $|\uparrow\downarrow\rangle$. The entanglement is revealed by the component of the parity signal that oscillates as $C_2\cos(2\phi_p)$, where $C_2 = |\rho_{\downarrow\downarrow,\uparrow\uparrow}|$. A value of $C_2 > 0.5$ verifies the spin entanglement of $|\Psi_f\rangle$ and thus the motional entanglement in state (5).

To deduce the spin populations, we use state-dependent resonance fluorescence[7,12]. The $|\uparrow\rangle$ state strongly fluoresces. Before measurement, we transfer the $|\downarrow\rangle$ population to a 'dark' auxiliary state (Methods). The populations $\varepsilon_{A,B}$ are in another dark auxiliary state, where they falsely contribute to $P_{\downarrow\downarrow}$ but in a way that does not depend on $\phi_p$. We fit the data in Fig. 3a with $C_2\cos(2\phi_p + \phi_2) + C_1\cos(\phi_p + \phi_1) + C_0$ and extract $C_2 = 0.57(2)$. This demonstrates that entanglement was present in the motion after the steps to create state (5).

The intermediate state (4) is itself a novel 'spin–motion' entangled state, where the spin state of the $^9\text{Be}^+$ ion in well B is entangled with the motion of the stretch mode of the ion pair in well A. We characterize this state in a separate set of experiments. After creating state (4), we allow it to persist for 176 μs. Following the analysis described above
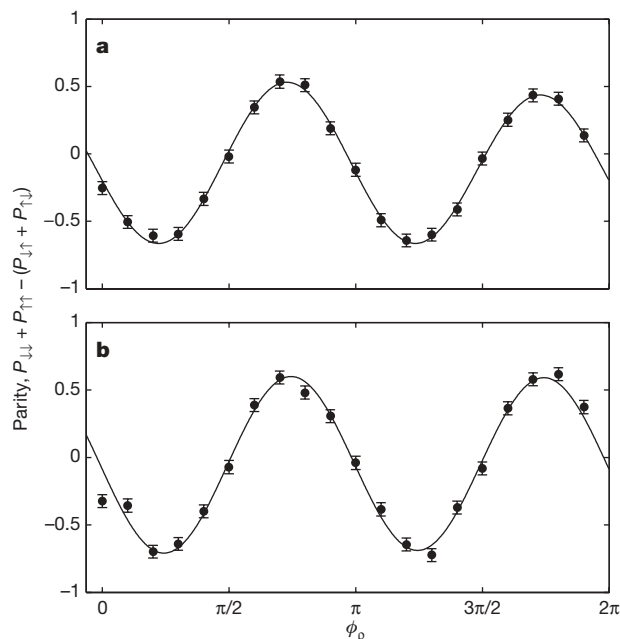
**Figure 3 | Entanglement demonstration through parity oscillation.** Parity data obtained from **a**, the entangled mechanical oscillators and **b**, the spin–motion entanglement experiments. Each point is calculated using the maximum-likelihood method on the fluorescence data from running the experiment 500 times, and is plotted with standard error bars (s.e.m.). Solid curves are fits to the data. Two-ion entanglement is verified by an amplitude greater than 0.5 for the component of the parity signal that oscillates at twice the analysis pulse phase, $\phi_p$ (ref. 26). For the data shown, this amplitude is **a**, 0.57(2) and **b**, 0.65(2).

(omitting the spin ↔ motion transfer steps in well B), we measure the parity (Fig. 3b) and find $C_2 = 0.65(2)$.

Significant sources of loss in fidelity are spontaneous photon scattering[27] and motional decoherence. The fidelity with which we initially create $|\Psi^+\rangle$ is ~0.88. Motional state superpositions of the stretch mode in each well were independently measured to have a coherence time of ~800 µs, which is consistent with a model of decoherence due to couplings to thermally occupied radial modes[28]. In the entangled mechanical oscillators experiment, the motional superpositions are occupied for ~250 µs and ~50 µs in wells A and B, respectively; we estimate a decrease in $C_2$ from this source to be ~5%. In the spin–motion entanglement experiment, we estimate a decrease in $C_2$ of ~3% from this source. Non-zero temperature also reduced the fidelity of motional state initialization. We estimate that this would reduce the fidelity for producing the states (5) and (4) by 8% and 6%, respectively. Intensity fluctuations at the few per cent level reduce the accuracy of all rotations.

The Coulomb coupling between the ion pairs in wells A and B could lead to an entangled state of their stretch modes. However, the resonant exchange rate is 5 Hz, which leads to negligible entanglement for the experimental timescales. Furthermore, the stretch mode frequencies in wells A and B differ by ~25 kHz, which would yield negligible entanglement for all timescales.

In summary, we have created two novel entangled states of separated systems involving mechanical oscillators, extending the regime where entanglement has been observed in nature. Implementing these experiments required deterministic ion ordering and the ability to separate and recool ions while preserving entanglement and performing subsequent coherent operations. This is, to our knowledge, the first demonstration of these techniques combined. Some of these methods could apply to similar experiments with nano- and micromechanical resonators[8–10]. The states created could be used to extend tests of non-locality in ion traps in a manner analogous to that proposed for the electromagnetic fields of separated cavities[11]. The control developed for

these experiments also represents an important step towards large-scale trapped-ion quantum information processing[12,14].

**METHODS SUMMARY**

The Methods section details (1) $^{24}Mg^+$ laser cooling of the motional modes of the two multi-species ion configurations, (2) protocols used for transferring population that was not mapped into the entangled motional state and transferring the final $|\downarrow\rangle$ population to a dark state, and (3) two control experiments. Supplementary Table 1 contains more details of the experimental sequence.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Schrödinger, E. Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften* **23**, 807–812,–823–828, 844–849 (1935).
2. Ball, P. Quantum all the way. *Nature* **453**, 22–25 (2008).
3. Schlosshauer, M. Lifting the fog from the north. *Nature* **453**, 39 (2008).
4. Bassi, A. & Ghirardi, G. Dynamical reduction models. *Phys. Rep.* **379**, 257–426 (2003).
5. Leggett, A. J. Testing the limits of quantum mechanics: Motivation, state of play, prospects. *J. Phys. Condens. Matter* **14**, R415–R451 (2002).
6. Aspelmeyer, M. & Zeilinger, A. A quantum renaissance. *Phys. World* **7**, 22–28 (2008).
7. Southwell, K. (ed.) Quantum coherence. *Nature* **453**, 1003–1049 (2008).
8. Mancini, S., Giovannetti, V., Vitali, D. & Tombesi, P. Entangling macroscopic oscillators exploiting radiation pressure. *Phys. Rev. Lett.* **88**, 120401 (2002).
9. Schwab, K. C. & Rourkes, M. L. Putting mechanics into quantum mechanics. *Phys. Today* **58**, 36–42 (2005).
10. Kippenberg, T. J. & Vahala, K. J. Cavity optomechanics: Back-action at the mesoscale. *Science* **321**, 1172–1176 (2008).
11. Milman, P. *et al.* A proposal to test Bell's inequalities with mesoscopic non-local states in cavity QED. *Eur. Phys. J. D* **32**, 233–239 (2005).
12. Wineland, D. J. *et al.* Experimental issues in coherent quantum-state manipulation of trapped atomic ions. *J. Res. Natl. Inst. Stand. Technol.* **103**, 259–328 (1998).
13. Cirac, J. I. & Zoller, P. A scalable quantum computer with ions in an array of microtraps. *Nature* **404**, 579–581 (2000).
14. Kielpinski, D., Monroe, C. & Wineland, D. Architecture for a large-scale ion-trap quantum computer. *Nature* **417**, 709–711 (2002).
15. Schleich, W. P. *Quantum Optics in Phase Space* 1st edn (Wiley-VCH, 2001).
16. Aspect, A. in *Quantum [Un]speakables — From Bell to Quantum Information* (eds Bertlmann, R. A. & Zeilinger, A.) Ch. 9 (Springer, 2002).
17. Matsukevich, D. N., Maunz, P., Moehring, D. L., Olmschenk, S. & Monroe, C. Bell inequality violation with two remote atomic qubits. *Phys. Rev. Lett.* **100**, 150404 (2008).
18. Pan, J.-W., Chen, Z.-B., Zukowski, M., Weinfurter, H. & Zeilinger, A. Multi-photon entanglement and interferometry. Preprint at ⟨http://arxiv.org/abs/0805.2853⟩ (2008).
19. Leibfried, D. *et al.* Experimental demonstration of a robust, high-fidelity geometrical two ion-qubit phase gate. *Nature* **422**, 412–415 (2003).
20. Rowe, M. A. *et al.* Transport of quantum states and separation of ions in a dual RF ion trap. *Quant. Inform. Comput* **2**, 257–271 (2002).
21. Barrett, M. D. *et al.* Deterministic quantum teleportation of atomic qubits. *Nature* **429**, 737–739 (2004).
22. King, B. E. *et al.* Cooling the collective motion of trapped ions to initialize a quantum register. *Phys. Rev. Lett.* **81**, 1525–1528 (1998).
23. Kielpinski, D. *et al.* A decoherence-free quantum memory using trapped ions. *Science* **291**, 1013–1015 (2001).
24. Barrett, M. D. *et al.* Sympathetic cooling of $^9Be^+$ and $^{24}Mg^+$ for quantum logic. *Phys. Rev. A* **68**, 042302 (2003).
25. Vandersypen, L. M. K. & Chuang, I. L. NMR techniques for quantum control and computation. *Rev. Mod. Phys.* **76**, 1037–1069 (2004).
26. Sackett, C. A. *et al.* Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
27. Ozeri, R. *et al.* Errors in trapped-ion quantum gates due to spontaneous photon scattering. *Phys. Rev. A* **75**, 042329 (2007).
28. Roos, C. F. *et al.* Nonlinear coupling of continuous variables at the single quantum level. *Phys. Rev. A* **77**, 040302(R) (2008).

685

## METHODS

**$^{24}$Mg$^+$ laser cooling of motional modes.** In addition to comprising part of the mechanical oscillators, the $^{24}$Mg$^+$ ions serve as a tool to provide sympathetic cooling of the $^9$Be$^+$ ions[24]. Doppler cooling of $^{24}$Mg$^+$ is accomplished by driving transitions between the ground $^2S_{1/2}$ states and excited $^2P_{1/2}$ states, which have a radiative linewidth of 41 MHz (ref. 29). In the 0.012 T applied magnetic field, the ground Zeeman states $|m_J = \pm1/2\rangle$ are split by 334 MHz, hence efficient Doppler cooling requires an additional repump beam to prevent optical pumping. One cycle of the pulsed $^{24}$Mg$^+$ sideband cooling[24] uses stimulated-Raman transitions on a motional sideband of the $|+1/2\rangle \rightarrow |-1/2\rangle$ ground state transition (duration $\sim$5 μs), followed by application of the repumping beam to reprepare $|+1/2\rangle$ ($\sim$2 μs).

At the start of each experiment, the four ions are located in one well in the configuration $^9$Be$^+$–$^{24}$Mg$^+$–$^{24}$Mg$^+$–$^9$Be$^+$, which has four axial modes of motion. In order of ascending frequency, these are the in-phase mode (frequency $\sim$2.0 MHz, mode vector: [0.32, 0.63, 0.63, 0.32]), the out-of-phase mode (4.1 MHz, [−0.47, −0.53, 0.53, 0.47]), a third mode (5.5 MHz, [0.63, −0.32, −0.32, 0.63]) and a fourth mode (5.7 MHz, [0.53, −0.47, 0.47, −0.53]). The amplitudes given in the mode vectors (written in ion order from left to right) are related to each ion's root-mean-squared ground state wavefunction size by multiplying by $\sqrt{\hbar/(2M\omega)}$, with $M$ the mass of the relevant ion and $\omega$ the mode frequency in angular units. There are also radial modes that have small amplitudes for $^9$Be$^+$ but large amplitudes for $^{24}$Mg$^+$. This means $^9$Be$^+$ cooling is inefficient for these modes, hence we also cool these modes using $^{24}$Mg$^+$ Doppler cooling. After preparing the four axial modes to near the ground state, the geometric phase gate[19] operation implements a $^9$Be$^+$ state-dependent motional displacement on the out-of-phase mode.

Sympathetic cooling plays a crucial role in making the transition from state (2) to state (3). After separating the ion pairs into wells A and B, we simultaneously cool them using $^{24}$Mg$^+$ Doppler cooling. This is followed by 40 cooling cycles per mode on the second motional sideband and then 60 cycles per mode on the first sideband to prepare the axial modes to near the ground state. The motional modes of each $^9$Be$^+$–$^{24}$Mg$^+$ pair are the 'common' mode (frequency $\sim$2.3 MHz, mode vector: [0.37, 0.93]) and the 'stretch' mode (4.9 MHz, [−0.93, 0.37]).

**Transfer to auxiliary hyperfine states.** Before final spin population measurement, the $|\downarrow\rangle$ population is transferred to the dark state $|F = 2, m_F = -2\rangle$ using carrier π pulses $R(\pi, 0)$ on the sequence of transitions $|2, 1\rangle \rightarrow |2, 0\rangle$, $|2, 0\rangle \rightarrow |2, -1\rangle$, $|2, -1\rangle \rightarrow |2, -2\rangle$. The number of photons measured per $^9$Be$^+$ ion if all the population were in the $|2, -2\rangle$ dark state during the 200 μs detection period approximates a Poisson distribution with a mean of 0.2. For the fluorescing state $|\uparrow\rangle$, we observe a Poisson distribution, with a mean number of photons $\sim$10 per $^9$Be$^+$ ion.

As described in the main text, we move populations $\varepsilon_{A,B}$ of the spin states $|\downarrow\rangle_{A,B}$ to the auxiliary hyperfine state $|2, 0\rangle$, so they do not contribute to entanglement verification. To ensure that the $\varepsilon_{A,B}$ populations end in dark states for the measurements, we precede the transfer pulses described in the previous paragraph with transfer of the $|2, 0\rangle$ populations to $|2, -2\rangle$ using a sequence of carrier π pulses on the $|2, 0\rangle \rightarrow |2, -1\rangle$ and $|2, -1\rangle \rightarrow |2, -2\rangle$ transitions. Since the last pulse of the final transfer sequence is also a carrier π pulse on the $|2, -2\rangle \leftrightarrow |2, -1\rangle$ transition, this leads to the populations $\varepsilon_{A,B}$ ending in $|2, -1\rangle$. If all the population is in this state, it would give a mean fluorescence value per $^9$Be$^+$ ion of $\sim$1 photon during detection. This fluorescence is independent of the final analysis pulse phase $\phi_p$, and hence does not contribute to $C_2$.

**Control experiments.** To provide partial checks of the spin → motion transfer steps, we perform separate experiments to determine the spin populations after the transfer. In the first check experiment, we follow the steps used to create state (4) then implement the above hyperfine state transfer sequences (omitting the $\varepsilon_B$ population transfer process) and measure the spin populations. The populations are determined to be $P_{\uparrow\uparrow} = 0.47(1)$, $P_{\downarrow\downarrow} = 0.04(1)$ and $P_{\downarrow\uparrow} + P_{\uparrow\downarrow} = 0.49(2)$. Ideally we would expect $P_{\uparrow\uparrow} = 1/2$, $P_{\downarrow\downarrow} = 0$ and $P_{\downarrow\uparrow} + P_{\uparrow\downarrow} = 1/2$. Similarly, after the step used to create state (5), and following the transfer procedure, we determine $P_{\uparrow\uparrow} = 0.86(2)$, $P_{\downarrow\downarrow} = 0.01(1)$ and $P_{\downarrow\uparrow} + P_{\uparrow\downarrow} = 0.13(2)$. Ideally we should find $P_{\uparrow\uparrow} = 1$.

29. Herrmann, M. *et al.* Frequency metrology on single trapped ions in the weak binding limit: The $3s_{1/2} - 3p_{3/2}$ transition in $^{24}$Mg$^+$. *Phys. Rev. Lett.* **102,** 013006 (2009).

# LETTERS

# Non-blinking semiconductor nanocrystals

Xiaoyong Wang[1], Xiaofan Ren[3], Keith Kahen[3], Megan A. Hahn[1], Manju Rajeswaran[3], Sara Maccagnano-Zacher[4], John Silcox[4], George E. Cragg[5], Alexander L. Efros[5] & Todd D. Krauss[1,2]

The photoluminescence from a variety of individual molecules[1] and nanometre-sized crystallites[2] is defined by large intensity fluctuations, known as 'blinking', whereby their photoluminescence turns 'on' and 'off' intermittently, even under continuous photoexcitation[2]. For semiconductor nanocrystals, it was originally proposed[3] that these 'off' periods corresponded to a nanocrystal with an extra charge. A charged nanocrystal could have its photoluminescence temporarily quenched owing to the high efficiency of non-radiative (for example, Auger) recombination processes between the extra charge and a subsequently excited electron–hole pair; photoluminescence would resume only after the nanocrystal becomes neutralized again. Despite over a decade of research, completely non-blinking nanocrystals[4,5] have not been synthesized and an understanding of the blinking phenomenon[6] remains elusive. Here we report ternary core/shell CdZnSe/ZnSe semiconductor nanocrystals that individually exhibit continuous, non-blinking photoluminescence. Unexpectedly, these nanocrystals strongly photoluminesce despite being charged, as indicated by a multi-peaked photoluminescence spectral shape and short lifetime. To model the unusual photoluminescence properties of the CdZnSe/ZnSe nanocrystals, we softened the abrupt confinement potential of a typical core/shell nanocrystal, suggesting that the structure is a radially graded alloy of CdZnSe into ZnSe. As photoluminescence blinking severely limits the usefulness of nanocrystals in applications requiring a continuous output of single photons, these non-blinking nanocrystals may enable substantial advances in fields ranging from single-molecule biological labelling[7] to low-threshold lasers[8].

Semiconductor nanocrystals have size-tunable optical properties that open up possibilities for revolutionary advances in lasers[8], light-emitting diodes[9], solar cells[10] and biological imaging[11–13]. However, the photoluminescence from individual nanocrystals, as well as from most other single molecular species (a notable exception is carbon nanotubes[14]), exhibits intermittent behaviour[1]. The existence of photoluminescence 'off' periods greatly limits the number of photons that can be detected in a given time period and also makes the photon arrival times from a single nanocrystal highly unpredictable. Thus, the synthesis of continuously emitting single nanocrystals would have profound influences on the usage of nanocrystals in applications for biology[7,11–13], quantum optics[15] and optoelectronics[8].

Recently, a significant effort has been devoted to understanding the underlying fundamental mechanisms for the photoluminescence blinking behaviour of single nanocrystals[3,6,16,17]. Although details are still unclear, a consensus has emerged that photoluminescence blinking is fundamentally caused by extra charges in a nanocrystal that greatly enhance non-radiative decay rates[3,6] (Fig. 1a). These charges arise from the separation of photoexcited electron–hole pairs, and are long-lived, as one charge carrier from the pair probably resides at the nanocrystal surface in deep-trap energy states. Thus one route to suppress blinking would be to eliminate trap sites. Indeed,

near-complete photoluminescence blinking suppression in single CdSe nanocrystals was recently achieved by adding surface bound ligands[18,19] and by growing thick semiconductor shells around the nanocrystal cores[4,5]. However, blinking was reduced, not eliminated, because the fundamental processes responsible for blinking were still present.

We have synthesized alloyed core nanocrystals of CdZnSe capped with a semiconductor shell of ZnSe (CdZnSe/ZnSe) that on the single nanocrystal level exhibit complete suppression of photoluminescence blinking on timescales from milliseconds to hours. Interestingly, these continuously emitting nanocrystals have a photoluminescence decay lifetime of about 5 ns, in contrast to the ~20 ns value routinely measured from traditional CdSe nanocrystals. Moreover, the single-nanocrystal photoluminescence spectra are highly unusual, displaying three distinct peaks. These unique and remarkable photophysical characteristics are well modelled by the radiative recombination of a trion (charged exciton; see below for details) in a nanocrystal in which non-radiative Auger processes are weakened.

Alloyed core CdZnSe nanocrystals were synthesized and capped with ZnSe using variations of literature procedures[20,21] (Supplementary Information). High resolution annular dark field scanning transmission electron microscopy showed that the nanocrystals are slightly prolate (Fig. 2a, Supplementary Fig. 1). Single-nanocrystal photoluminescence measurements were acquired on a confocal microscope (Fig. 2b). Significantly, the photoluminescence from single nanocrystals was continuously 'on' with no fluctuations or 'off' periods on timescales from milliseconds to hours (Fig. 2c, Supplementary Fig. 2). This continuous emission is in stark contrast
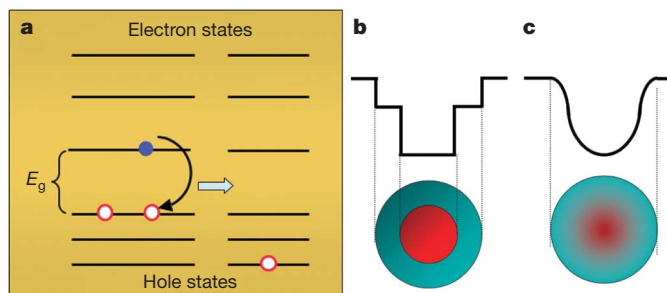


**Figure 1 | Non-radiative Auger-like recombination and potential energy functions of nanocrystals. a**, Diagram of an Auger process. $E_g$ corresponds to the nanocrystal energy gap. An electron (filled blue circle) non-radiatively recombines with a hole (red circle), producing an electronically excited hole. **b**, The potential energy function presents a series of steps for traditional nanocrystals. The undefined slope at the boundary between unlike materials results in an uncertainty in momentum. **c**, A gradually changing potential energy function exists for non-blinking CdZnSe/ZnSe nanocrystals.

[1]Department of Chemistry, [2]The Institute of Optics, University of Rochester, Rochester, New York 14627, USA. [3]Eastman Kodak Company, Rochester, New York 14650, USA. [4]School of Applied and Engineering Physics, Cornell University, Ithaca, New York 14583, USA. [5]Naval Research Laboratory, Washington DC 20375, USA.
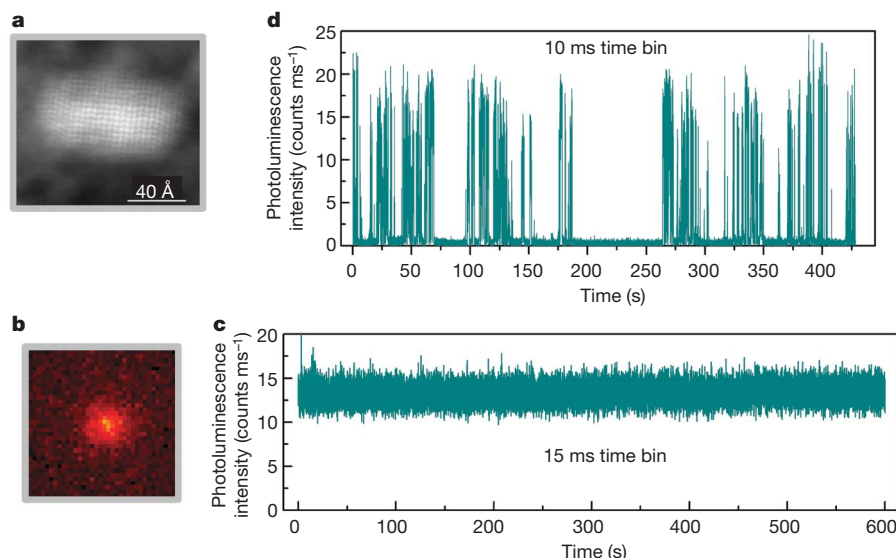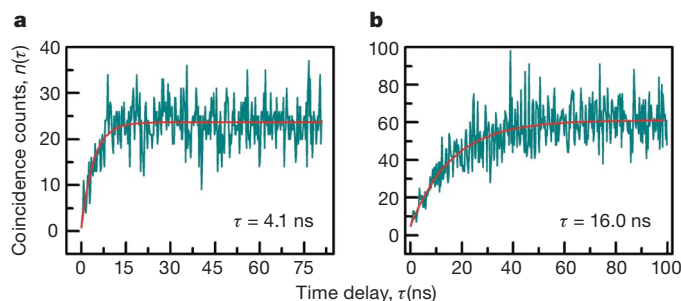
**Figure 2 | Single-nanocrystal images and photoluminescence intensity time traces. a**, Electron micrograph of a CdZnSe/ZnSe nanocrystal with diameter ~5 nm and length ~6.7 nm. Chemical information with atomic resolution was not achieved owing to difficulties obtaining energy-loss signals from Zn, and the weak contrast between Cd and Zn (ref. 27).

**b**, Photoluminescence image (~3 μm × 3 μm) from a single CdZnSe/ZnSe nanocrystal (different nanocrystal from **a**). **c, d**, Time dependent photoluminescence intensity traces from a single CdZnSe/ZnSe nanocrystal (**c**) and a CdSe/ZnS nanocrystal (**d**). Similar time traces were obtained from 118 (>98%) of 120 CdZnSe/ZnSe nanocrystals studied.

to a typical CdSe/ZnS core/shell nanocrystal, which displays photoluminescence blinking (Fig. 2d).

To determine that the photoluminescence was indeed from a single nanocrystal, we performed photon anti-bunching measurements (Fig. 3a). The dip observed around zero time delay in the coincidence counts unambiguously confirms that a single nanocrystal is being investigated[15]. The anti-bunching curve can be fitted with a single exponential growth function, from which a photoluminescence lifetime of ~4.1 ns can be extracted, which is significantly shorter than the ~20 ns photoluminescence lifetime measured from CdSe/ZnS nanocrystals that exhibit blinking[15,22] (Fig. 3b).

Figure 4a shows the photoluminescence spectra of five selected single CdZnSe/ZnSe nanocrystals, which contain multiple emission peaks that span almost the entire wavelength range of the ensemble photoluminescence spectrum (Supplementary Fig. 3). The single-nanocrystal spectrum is quite different from the symmetric lineshape of a typical single CdSe nanocrystal, shown in Fig. 4b. Interestingly, although the absolute energies of the photoluminescence peaks depend strongly on the particular single nanocrystal being studied (Fig. 4a), the energy difference between the first (second) and second (third) peaks is relatively constant at ~164 meV (Supplementary Fig. 4, Supplementary Table 1).

An appropriate model of the photoluminescence properties of CdZnSe–ZnSe nanocrystals must account for (1) the lack of photoluminescence blinking, (2) the equidistant multiple peaks in the photoluminescence spectrum, and (3) the relatively short photoluminescence lifetime. We considered the recombination of a charged exciton (that is, a trion, formed from a photogenerated exciton and a long-lived charge in a nominally ionized nanocrystal) confined in a parabolic, but axially symmetric potential, including hole–hole and electron–hole interactions (Supplementary Information page 6), which can simultaneously account for the three exceptional features of the CdZnSe/ZnSe nanocrystal photoluminescence. The electron and hole energy spectra consist of equidistant discrete levels:

$$E_{e,h} = \hbar\omega_z^{e,h}(n_z^{e,h} + 1/2) + \hbar\omega_\perp^{e,h}(n_x^{e,h} + 1/2) + \hbar\omega_\perp^{e,h}(n_y^{e,h} + 1/2) \quad (1)$$

where $\omega_z^{e,h}$ ($\omega_\perp^{e,h}$) are characteristic frequencies for electrons and holes along the elongated z-axis of the nanocrystal (or perpendicular to this axis), and $n_{\{x,y,z\}}$ are the respective quantum numbers. We determined that the ground singlet state of the trion consists of two holes with opposite spins and an electron occupying the lowest conduction band level with $n_{z,x,y}^{e,h} = 0$ (Supplementary Information). Positively charged nanocrystals are consistent with previous studies on photoexcited CdSe nanocrystals[23].

Importantly, after radiative recombination the remaining hole can populate one of several equally spaced energy levels (with quantum numbers $n_x$, $n_y$, $n_z$ of equation (1)), and thus multiple peaks in the photoluminescence lineshape are expected (Fig. 4a). Interestingly, photoluminescence spectra from charged excitons in semiconductors also display multiple peaks, called 'shake-up lines'[24], whose energies are determined by the difference between the trion annihilation energy and an excitation energy of the extra hole in its final state (Fig. 4c). The peaks in the photoluminescence spectrum are clearly electronic in origin, and originate from the same excited state, as emission from multiple excitons within the nanocrystal would not cause a dip in the anti-bunching signal. Vibrational contributions are excluded, as the energy difference between adjacent peaks is much larger than the $kT$ energy (~23 meV) as well as being larger than the longitudinal-optical phonon energies of ~25 meV and ~31 meV for CdSe and ZnSe, respectively[25]. Finally, spectral shifting[26] is also excluded owing to the magnitude of the peak separation as well as the observation of discrete separation energies.



**Figure 3 | Anti-bunching measurements. a**, Anti-bunching curve of a typical single CdZnSe/ZnSe nanocrystal. **b**, Anti-bunching curve of a standard CdSe/ZnS nanocrystal. The solid lines are fits with single exponential growth functions with corresponding lifetimes of 4.1 ns and 16.0 ns, respectively. Similar anti-bunching results were obtained from all other single CdZnSe/ZnSe nanocrystals studied (40 nanocrystals in total) and photoluminescence lifetimes were in the range 4.0–5.6 ns.

**Figure 4 | Single-nanocrystal photoluminescence spectra and the shake-up process. a**, Photoluminescence spectra of five selected single CdZnSe/ZnSe nanocrystals. **b**, Photoluminescence spectrum of a single CdSe/ZnS nanocrystal. **c**, Diagram of a shake-up process used to explain the multi-peaked photoluminescence spectrum from a single CdZnSe/ZnSe nanocrystal. The annihilation energy of the trion (spacing between two dashed lines) is distributed between the emitted photon energy (shown by blue, green and red arrows) and the energy of the extra hole, which could occupy one of many allowed levels after recombination.

Although the exact form of the parabolic potential is unknown, we found reasonable parameters describing the potential that are consistent with the energy difference between the photoluminescence peaks ($\sim$164 meV) and with the ratios of the integrated photoluminescence intensities of the second and third peaks to the first (0.87 and 0.26, respectively; Supplementary Information). The photoluminescence spectrum is sensitive to parameters describing the localization volume of the hole and is much less sensitive to the electron-related parameters. Minimization of the trion binding energy with respect to these parameters produces a potential that strongly localizes the holes at the nanocrystal centre, while the electron wavefunction spreads across the entire nanocrystal volume, as expected.

Soft confinement of electrons and holes is critical for suppression of Auger recombination, which is important because most theories of nanocrystal blinking[6] are refinements to the original Auger ionization model[3]. Efficient Auger recombination requires breaking strict momentum conservation, such as in typical core–shell nanocrystals where the change from one material to another is extremely abrupt[27], causing the potential to resemble a series of sharp steps (Fig. 1b). Our model suggests CdZnSe/ZnSe nanocrystals have a gradually changing confinement potential (Fig. 1c). Thus, the rate of the non-radiative recombination is expected to be mitigated[28], which is consistent with trion emission, as a trion should not emit light in nanocrystals with unsuppressed Auger processes[3].

In typical CdSe nanocrystals, the long photoluminescence lifetime ($\sim$20 ns at 300 K) is connected with the lowest excited state having a forbidden optical transition dipole (in other words, it is a 'dark state')[29]. However, radiative recombination of singlet trions is always optically allowed, because one of the hole spins (either spin up or down) forming the singlet can always recombine with any electron spin. We calculated the trion radiative decay time to be between 8 and 10 ns (Supplementary Information), in reasonable agreement with the experimentally determined photoluminescence lifetime (given 50% ensemble quantum efficiency). Although the calculation of the lifetime provides some validation of the soft-confinement model,

it would also be reasonably consistent with the lifetime of typical CdSe nanocrystals, given the uncertainties in the calculation. Assuming that the dominant non-radiative processes of charged trions are due to Auger-like processes, we estimate (for a nanocrystal with a 50% quantum efficiency) a lower bound on the Auger decay time that is also around 10 ns, implying one order of magnitude suppression in this rate for CdZnSe/ZnSe nanocrystals.

The collective CdZnSe/ZnSe photoluminescence data are best explained by the radiative recombination of a positive trion in a nanocrystal with a soft-confinement potential. Nanocrystal absorption, being well above the band edge, is not affected by the extra hole, while the soft-confinement potential, which probably arises from a graded alloy CdZnSe core, reduces the efficiency of Auger processes so that a nanocrystal is continuously emitting (that is, always 'on'), whether it is charged or neutral. Although other theories (such as an elimination of surface state traps) can explain the lack of photoluminescence blinking, they can not explain the photoluminescence spectrum (they would predict a single Lorentzian line) and in fact predict longer (not shorter) photoluminescence lifetimes.

Interestingly, some self-organized quantum dots (for example, epitaxially grown InGaAs on GaAs) also have photoluminescence from trions and bi-excitons[30] but with little photoluminescence blinking, owing to smooth potential energy functions attributed to graded interfaces between the materials comprising the quantum dot. Colloidal soft-confinement alloyed-core CdZnSe/ZnSe nanocrystals approach the quality of self-organized quantum dots, but with the added benefit of having cheap and facile wet-chemical processability. As such, non-blinking nanocrystals may enable significant advances in single-molecule biological labelling and tracking, as their continuous and bright emission makes for easier fluorescence detection and identification[7]. Because Auger processes are primarily responsible for the difficulty in achieving optical gain from colloidal nanocrystals[8], non-blinking nanocrystals should have significantly reduced lasing thresholds, allowing for the exciting possibility of electrical (as opposed to intense optical) pumping.

CdZnSe/ZnSe nanocrystals are not without limitations; for example, the broad photoluminescence lineshape will make it difficult to simultaneously track multiple biomolecules, or to have significant spectral multiplexing in nanocrystal optical amplifiers. In addition, technical development issues such as biocompatibility or incorporation into optical gain media must be demonstrated before applications can be pursued. However, the path to pioneering technology is clear, as similar challenges have been solved for CdSe nanocrystals[8,11–13], and it is unlikely that the unique properties of CdZnSe/ZnSe nanocrystals will be affected by subsequent processing.

## METHODS SUMMARY

Single nanocrystal samples were prepared by spin-casting a dilute solution of CdZnSe/ZnSe nanocrystals (synthesized at Kodak) in toluene or CdSe/ZnS nanocrystals in water onto a quartz coverslip. Nanocrystals placed on a confocal optical microscope were excited at 532 nm through an oil immersion objective (NA = 1.3) at 300 K with a power density of $\sim 1\,kW\,cm^{-2}$. Photoluminescence was collected by the same objective and sent to a charge coupled device (CCD) attached to a spectrometer for photoluminescence spectra and images, or to a time-correlated single-photon counting system for the blinking and anti-bunching measurements. Photon count rates are consistent with an excitation rate of $\sim 1\,MHz$.

1.  Moerner, W. E. & Orrit, M. Illuminating single molecules in condensed matter. *Science* **283**, 1670–1676 (1999).
2.  Nirmal, M. *et al.* Fluorescence intermittency in single cadmium selenide nanocrystals. *Nature* **383**, 802–804 (1996).
3.  Efros, Al. L. & Rosen, M. Random telegraph signal in the photoluminescence intensity of a single quantum dot. *Phys. Rev. Lett.* **78**, 1110–1113 (1997).
4.  Chen, Y. F. *et al.* "Giant" multishell CdSe nanocrystal quantum dots with suppressed blinking. *J. Am. Chem. Soc.* **130**, 5026–5027 (2008).
5.  Mahler, B. *et al.* Towards non-blinking colloidal quantum dots. *Nature Mater.* **7**, 659–664 (2008).
6.  Frantsuzov, P., Kuno, M., Jankó, B. & Marcus, R. A. Universal emission intermittency in quantum dots, nanorods and nanowires. *Nature Phys.* **4**, 519–522 (2008).
7.  Dahan, M. *et al.* Diffusion dynamics of glycine receptors revealed by single-quantum dot tracking. *Science* **302**, 442–445 (2003).
8.  Klimov, V. I. *et al.* Optical gain and stimulated emission in nanocrystal quantum dots. *Science* **290**, 314–317 (2000).
9.  Colvin, V. L., Schlamp, M. C. & Alivisatos, A. P. Light-emitting-diodes made from cadmium selenide nanocrystals and a semiconductor polymer. *Nature* **370**, 354–357 (1994).
10. Nozik, A. J. Spectroscopy and hot electron relaxation dynamics in semiconductor quantum wells and quantum dots. *Annu. Rev. Phys. Chem.* **52**, 193–231 (2001).
11. Bruchez, M. Jr, Moronne, M., Gin, P., Weiss, S. & Alivisatos, A. P. Semiconductor nanocrystals as fluorescent biological labels. *Science* **281**, 2013–2016 (1998).
12. Chan, W. C. W. & Nie, S. Quantum dot bioconjugates for ultrasensitive nonisotopic detection. *Science* **281**, 2016–2018 (1998).
13. Michalet, X. *et al.* Quantum dots for live cells, *in vivo* imaging, and diagnostics. *Science* **307**, 538–544 (2005).
14. Hartschuh, A., Pedrosa, H. N., Novotny, L. & Krauss, T. D. Simultaneous fluorescence and Raman scattering from single carbon nanotubes. *Science* **301**, 1354–1356 (2003).
15. Michler, P. *et al.* Quantum correlation among photons from a single quantum dot at room temperature. *Nature* **406**, 968–970 (2000).
16. Kuno, M., Fromm, D. P., Hamann, H. F., Gallagher, A. & Nesbitt, D. J. "On"/"off" fluorescence intermittency of single semiconductor quantum dots. *J. Chem. Phys.* **115**, 1028–1040 (2001).
17. Shimizu, K. T. *et al.* Blinking statistics in single semiconductor nanocrystal quantum dots. *Phys. Rev. B* **63**, 205316 (2001).
18. Hohng, S. & Ha, T. Near-complete suppression of quantum dot blinking in ambient conditions. *J. Am. Chem. Soc.* **126**, 1324–1325 (2004).
19. Fomenko, V. & Nesbitt, D. J. Solution control of radiative and nonradiative lifetimes: a novel contribution to quantum dot blinking suppression. *Nano Lett.* **8**, 287–293 (2008).
20. Talapin, D. V. *et al.* CdSe/CdS/ZnS and CdSe/ZnSe/ZnS core-shell-shell nanocrystals. *J. Phys. Chem. B* **108**, 18826–18831 (2004).
21. Zhong, X., Han, M., Dong, Z., White, T. J. & Knoll, W. Composition-tunable $Zn_xCd_{1-x}Se$ nanocrystals with high luminescence and stability. *J. Am. Chem. Soc.* **125**, 8589–8594 (2003).
22. Crooker, S. A., Hollingsworth, J. A., Tretiak, S. & Klimov, V. I. Spectrally resolved dynamics of energy transfer in quantum-dot assemblies: towards engineered energy flows in artificial materials. *Phys. Rev. Lett.* **89**, 186802 (2002).
23. Krauss, T. D. & Brus, L. E. Charge, polarizability, and photoionization of single semiconductor nanocrystals. *Phys. Rev. Lett.* **83**, 4840–4843 (1999).
24. Paskov, P. P. *et al.* Auger processes in InAs self-assembled quantum dots. *Physica E* **6**, 440–443 (2000).
25. Hellwege, K. H. (ed.) *Semiconductors. Physics of II–VI and I–VII Compounds, Semimagnetic Semiconductors*, Landolt-Börnstein, New Series, Group III, Vol. 17, Pt b (Springer, 1983).
26. Empedocles, S. A., Norris, D. J. & Bawendi, M. G. Photoluminescence spectroscopy of single CdSe nanocrystallite quantum dots. *Phys. Rev. Lett.* **77**, 3873–3876 (1996).
27. Yu, Z., Guo, L., Du, H., Krauss, T. D. & Silcox, J. Shell distribution on colloidal CdSe/ZnS quantum dots. *Nano Lett.* **5**, 565–570 (2005).
28. Landsberg, P. T. *Recombination in Semiconductors* (Cambridge Univ. Press, 1991).
29. Efros, Al. L. *et al.* Band-edge exciton in quantum dots of semiconductors with a degenerate valence band: dark and bright exciton states. *Phys. Rev. B* **54**, 4843–4856 (1996).
30. Dekel, E. *et al.* Multiexciton spectroscopy of a single self-assembled quantum dot. *Phys. Rev. Lett.* **80**, 4991–4994 (1998).

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.D.K. (krauss@chem.rochester.edu), A.L.E. (efros@dave.nrl.navy.mil) or K.K. (keith.kahen@kodak.com).

# LETTERS

# The Gamburtsev mountains and the origin and early evolution of the Antarctic Ice Sheet

Sun Bo[1], Martin J. Siegert[2], Simon M. Mudd[2], David Sugden[2], Shuji Fujita[3], Cui Xiangbin[1], Jiang Yunyun[1], Tang Xueyuan[1] & Li Yuansheng[1]

Ice-sheet development in Antarctica was a result of significant and rapid global climate change about 34 million years ago[1]. Ice-sheet and climate modelling suggest reductions in atmospheric carbon dioxide (less than three times the pre-industrial level of 280 parts per million by volume) that, in conjunction with the development of the Antarctic Circumpolar Current, led to cooling and glaciation paced by changes in Earth's orbit[2]. Based on the present subglacial topography, numerical models point to ice-sheet genesis on mountain massifs of Antarctica, including the Gamburtsev mountains at Dome A, the centre of the present ice sheet[2,3]. Our lack of knowledge of the present-day topography of the Gamburtsev mountains[4] means, however, that the nature of early glaciation and subsequent development of a continental-sized ice sheet are uncertain. Here we present radar information about the base of the ice at Dome A, revealing classic Alpine topography with pre-existing river valleys overdeepened by valley glaciers formed when the mean summer surface temperature was around 3 °C. This landscape is likely to have developed during the initial phases of Antarctic glaciation. According to Antarctic climate history (estimated from offshore sediment records) the Gamburtsev mountains are probably older than 34 million years and were the main centre for ice-sheet growth. Moreover, the landscape has most probably been preserved beneath the present ice sheet for around 14 million years.

Deep-sea oxygen isotope records show that the Eocene and Oligocene epochs represent times of global cooling culminating in the development of the first Antarctic Ice Sheet and an important expansion of Antarctic ice volume[1]. The Eocene (~52 to ~34 million years (Myr) ago) is characterized by a global cooling trend which continued during the remainder of the Cenozoic era. Subsequently there were two stepped changes in the rate of cooling. The first, at the Eocene–Oligocene boundary ~34 Myr ago, saw the onset of significant glaciation in Antarctica. The second, at ~14 Myr ago, is recorded by a 6–7 °C cooling in the marine isotope record[5,6] and in terrestrial evidence of cooling of at least 8 °C in the Transantarctic mountains[7].

Two approaches to modelling the initial growth of the Antarctic Ice Sheet show that glaciation begins in the upland mountain massifs of Antarctica, at coastal Dronning Maud Land, the Transantarctic mountains, and the Gamburtsev mountains beneath Dome A[2,8]. This central dome dominates glaciation because of its high altitude and consequent cold surface temperatures. Ice-sheet modelling, ocean cores and stratigraphic evidence suggest that for 20 million years, from 34 to 14 Myr ago, Antarctica experienced orbitally driven ice-volume fluctuations similar in scale to those of the Pleistocene ice sheets of the Northern Hemisphere and that these fluctuations were accompanied by marked changes in global sea level[2,9–11]. Tundra

biota survived at high altitudes during this period[7]. After 14 Myr the ice sheet, at least in higher mountain peripheries in East Antarctica, maintained its presence and control over the cold polar climate of today, leading to extremely low rates of erosion[12], cold-based local glaciers[13] and even the preservation of buried Miocene ice[14].

Our knowledge of the subglacial topography at Dome A has been obtained during only one radar flight in the 1970s[4,15,16]. Consequently, the present form and evolution of the Gamburtsev mountains are poorly understood, making models of ice-sheet inception problematic. Indeed, the morphology of the mountains is less well known than the surface of Mars.

In seasons 2004/05 and 2007/08, Chinese glaciologists made the first detailed radar survey of the Gamburtsev mountains (as part of the International Polar Year programme Chinese Antarctic Research Expedition; CHINARE). The bed was detected in the majority of radar lines (Fig. 1), and by subtracting ice thickness from surface elevation (measured by GPS) the elevation of the bed could be found. The bed elevations were then interpolated[17] onto a regular grid with pixel resolution of 140.5 m (see Methods Summary and Supplementary Methods for interpolation details). The unprecedented density of radar transects in this region means that the resulting Digital Elevation Model (DEM) provides the first detailed depiction of the topography of the central Gamburtsev mountains (Fig. 2).

The topography revealed beneath the ice is striking (Fig. 2 and Supplementary Fig. 1). The region consists of a south-facing elongated valley head, cutting over a kilometre into flanking mountains. The whole region is covered by ice 1,649–3,135 m thick. The maximum elevation of the topography is 2,434 m above sea level at 80° 18′ S, 76° 10′ E. The valley geometry is dendritic. We highlight this geometry by extracting a drainage network using standard methods[18] (Fig. 2, Supplementary Discussion 1). Recent numerical modelling, backed by empirical observations, has shown that ice cannot create such networks alone; subglacial topography takes this form only when ice exploits pre-existing fluvial topography (Supplementary Fig. 2)[8,19]. This fluvial landscape has subsequently been subject to intense valley glaciation, as demonstrated by overdeepening in the valley floors of up to 432 m and the presence of steep trough sides. It is also shown by details such as the location of overdeepened basins at points of valley convergence, staircases of intervening riegels or valley steps, hanging tributary valleys, and corries with steep arcuate cliffs and flat floors at the head of some tributary valleys (Fig. 3); such features are characteristic of landscapes shaped by valley glaciers[20,21]. Hanging valleys are formed when ice ponds in tributary glaciers as they enter the trunk glacier; this ponding leads to reduced ice surface slopes, which in turn reduces shear stress and sliding velocities at the glacier bed, ultimately reducing erosive capacity in the tributary glacier[20,21]. Another effect of
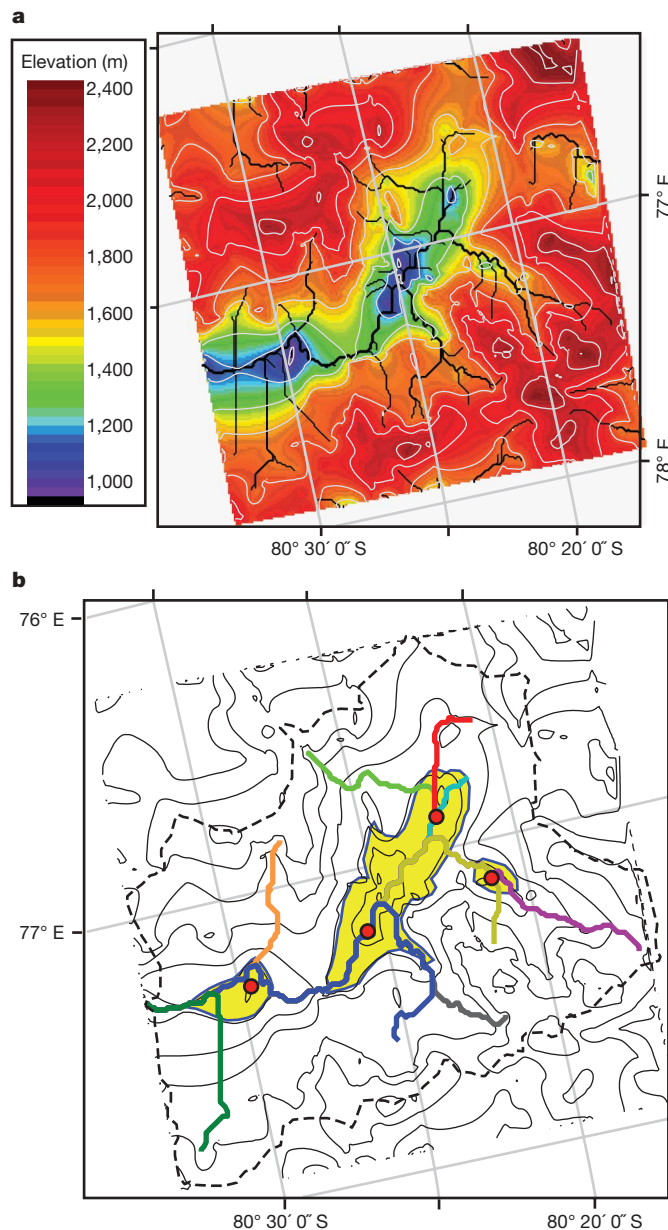
[1]Polar Research Institute of China, 451 Jinqiao Road, Pudong, Shanghai, 200136, China. [2]School of GeoSciences, University of Edinburgh, King's Buildings, Edinburgh EH9 3JW, UK. [3]National Institute of Polar Research, Research Organization of Information and Systems, Kaga, 1-9-10, Itabashi-ku, Tokyo 173-8515, Japan.
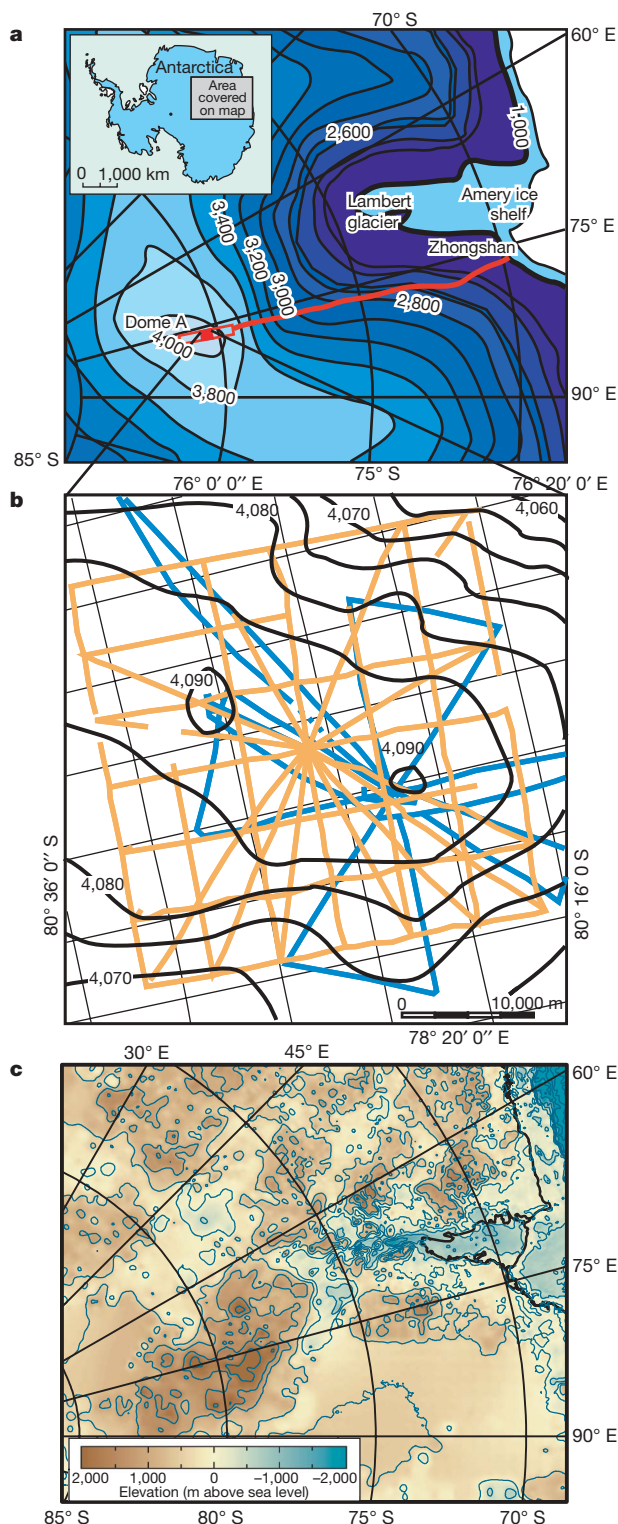
**Figure 1 | Location of Dome A and orientation of radar transects.**
**a,** Location of the 30 km × 30 km radar network. Ice sheet surface contours are provided at intervals of 200 m above 2,000 m, and at 1000 m. **b,** Radar data acquired in 2004/5 are shown in blue and from 2007/8 in orange. The area covered in the map is the same as in Fig. 2b. Ice sheet surface contours are provided in metres above sea level. Ice flow is perpendicular to these contours. **c,** Existing bed depiction of the Dome A region[4]. Elevation (metres above sea level) is provided in the scale bar.



**Figure 2 | Subglacial topography of the central Gamburtsev mountains.**
**a,** Topography of the surveyed section of the Gamburtsev mountains. Topographic contours (spaced at 200 m, lowest contour at 1,000 m) in grey; extracted valley network in black. Local elevation minima (red dots in **b**) are those which lie more than 150 m below the nearest down-valley saddle; this elevation change is well in excess of the vertical error (see main text). **b,** Selected features of the topography. Overdeepened areas are shown in yellow. Coloured channels are those selected for longitudinal profile analysis (Fig. 3). Main basin (562 km²) delimited with black dotted line. Contours provided as in **a**. Present day ice flow is to the North.

a tributary entering a trunk glacier is an increase in ice discharge in the trunk glacier, which increases the sliding rate and causes overdeepening[20,21]. The landscape has relief similar to the European Alps, yet the largest valley revealed by our survey is broader than valleys of similar drainage area in the Alps and other glaciated terrains such as the Olympic peninsula in Washington and central Idaho, both in the United States[22,23] (Supplementary Discussion 2). We note that the landscape identified beneath Dome A is different to that measured beneath other East Antarctic ice-sheet divides, such as at Dome C where the relief is far more subdued and low-lying[24].

Three phases of glaciation are apparent. First, small mountain-top glaciers occupied the corries, causing the development of discrete cliffed basins. Second, an expanded valley glacier occupied the central trough, fed by ice accumulating either on the surrounding mountains or on ice caps centred over the mountains. This glacier flowed south,
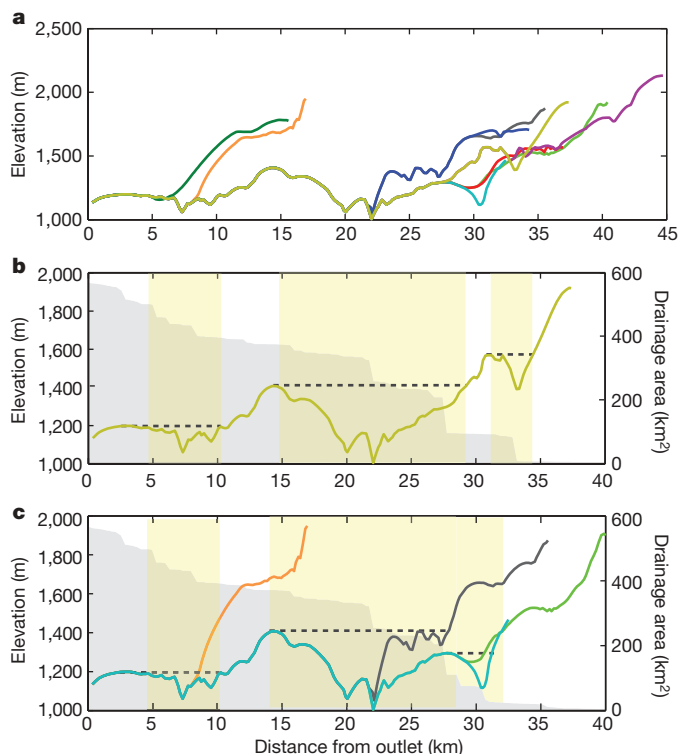
**Figure 3 | Longitudinal profiles and drainage area–distance plots of Dome A tributary valleys. a**, Profiles of selected valleys (colouring as in Fig. 2b). Several tributaries have steep headwaters indicative of corrie excavation. Hanging valleys and overdeepenings downstream of confluences are also visible. Overdeepenings are more than 150 m below the closest down valley bar. **b**, Profile (yellow) and drainage area (grey) of an individual valley, highlighting overdeepenings (dashed lines) corresponding to sections of valley where drainage area increases sharply (owing to tributary confluences). **c**, Further tributary profiles and drainage areas. The grey drainage area–distance plot is for the tributary marked in green.

and eroded the valley floor and sidewalls to their present shape, leading to a series of hanging valleys. Third, there is the modern ice sheet, which subsumes the entire region with cold slow-flowing ice. None of the landforms are conformable with present-day ice flow (Fig. 1b).

The three glacial phases have very different climatic needs. Today's ice-surface climate involves hyper-cold conditions (likely to be the coldest on the planet's surface) and no melting. Under such conditions, glaciological theory shows that the basal ice remains below the pressure melting point and is frozen to the bed, thus preventing significant glacial erosion and preserving the topography beneath[25]. For the two smaller-scale glacial phases, the climate must have been different. Both require an equilibrium line altitude separating glacier accumulation and ablation zones that is higher than the surrounding lowlands. Allowing for isostatic recovery of 500–1,000 m in view of the absence of the ~3-km-thick East Antarctic Ice Sheet at the onset of glaciation[8,24] (see Supplementary Discussion 3), the landscape would have been at an altitude of 1,500–3,500 m in pre-glacial times. This must imply an equilibrium line altitude of ~1,500–2,000 m above present sea level. Clearly, the larger valley glaciers would exist in colder periods, while corrie glaciers would exist when the climate was warmer. Thermomechanical glaciological models show that local Antarctic glaciers in the Transantarctic mountains would need maximum summer temperatures of the order of 3 °C or above if they are to have warm basal ice[7]. The implication is that summer temperatures in central East Antarctica at altitudes above 1,500–2,000 m in the early stages of mountain glaciation would have been no lower than about 3 °C.

The scale of the warming represented by the glacial landforms in the Gamburtsev mountains is unlikely to have occurred in Antarctica

for the past 14 Myr (refs 5 and 8). Hence, we believe it is likely that the glacial topography observed predates this time. The dendritic valley networks probably date back to the topography of Antarctica achieved in the Eocene when there were beech trees in coastal Antarctica and soils containing the clay mineral smectite typical of forests[26,27]. Alpine glaciation of the Gamburtsev mountains must have began at least by 34 Myr ago and may have continued during several intervals during the next 20 million years. One implication is that the Gamburtsev mountains predate 34 Myr ago, although at this stage we cannot rule out synchronous uplift and glaciation around this time. If the orbitally forced glacial changes were on the scale of the Northern Hemisphere fluctuations of the Pleistocene epoch, then there would also have been many occasions when the mountains lay buried beneath larger and colder ice sheets that protected the mountains beneath cold-based ice. In such a case the Alpine glaciation could have marked warmer periods. What seems more certain is that the mountains and their record of early warm-based glaciation have remained buried and preserved by the present cold-based ice sheet, which nucleated at the Gamburtsev mountains and grew to continental size[28], for the past 14 Myr. The alternative view, that the upland glacial landforms could have been formed in the Pliocene epoch[29], requires the loss of the bulk of the Antarctic Ice Sheet, with significant implications for global temperature and sea level that are not observed in proxy records at this time[1].

## METHODS SUMMARY

Radar information, collected by overland traversing at a frequency of 179 MHz, was gathered across a 30 × 30 km section of Dome A (Fig. 1). The two-way travel times of radio reflections from the ice-sheet base were converted to ice thickness, thus revealing information about the subglacial relief. The accuracy of the radar was assessed in 111 transect crossover points, revealing errors of <50 m in 64% of them, which in high-relief terrain is excellent (Supplementary Fig. 2). Around the centre of the survey region the transect separation is less than 5 km. To build a surface DEM from a series of transects requires us to interpolate between data points. Several interpolation procedures are commonly available, but we chose a technique called 'natural neighbour' interpolation, because it is designed to be used with distributed data confined to discrete transects[17]. For such data arrangements, this technique has been shown to give better results than interpolation algorithms designed for use on randomly distributed data[30], which can lead to artefacts introduced in the DEM. To examine how alternative data interpolation procedures may yield differences in DEMs, we processed the radar data using a variety of algorithms (see Supplementary Methods). Although the DEM results did differ slightly, the main topographic features discussed in this paper were consistently reproduced by all the interpolation techniques used. Importantly, the main valleys and depressions discussed are each covered by radar transects and, therefore, several data points (given the along-track data point separation of ~125 m). We are therefore confident that the DEM presented in Fig. 2a is not subject to artefacts that adversely affect the conclusions of this paper.

1. Zachos, J. C., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292, 686–693 (2001).
2. DeConto, R. M. & Pollard, D. Rapid Cenozoic glaciation of Antarctica induced by declining atmospheric CO$_2$. *Nature* 421, 245–249 (2003).
3. Huybrechts, P. Glaciological modelling of the Late Cenozoic East Antarctic ice sheet: stability or dynamism? *Geogr. Ann.* 75, 221–238 (1993).
4. Lythe, M., Vaughan, D. G., &. BEDMAP Consortium. BEDMAP: a new thickness and subglacial topographic model of Antarctica. *J. Geophys. Res.* 106, 11335–11352 (2001).
5. Shevenell, A. E., Kennett, J. P. & Lea, D. W. Middle Miocene Southern Ocean cooling and Antarctic cryosphere expansion. *Science* 305, 1766–1770 (2004).
6. Holbourn, A., Kuhn, W., Schulz, M. & Erlenkeuser, H. Impacts of orbital forcing and atmospheric carbon dioxide on Miocene ice-sheet expansion. *Nature* 438, 483–487 (2005).
7. Lewis, A. R. et al. Mid-Miocene cooling and the extinction of tundra in continental Antarctica. *Proc. Natl Acad. Sci. USA* 105, 10676–10680 (2008).
8. Jamieson, S. S. R. & Sugden, D. E. in *Antarctica, a Keystone in a Changing World* (eds Cooper, A. et al.) 39–54 (National Academies Press, 2007).
9. Pekar, S. F. & DeConto, R. M. High-resolution ice-volume estimates for the early Miocene: evidence for a dynamic ice sheet in Antarctica. *Palaeogeog. Palaeoclimatol. Palaeoecol.* 231, 101–109 (2005).
10. Naish, T. R. et al. Orbitally induced oscillations in the East Antarctic ice sheet at the Oligocene/Miocene boundary. *Nature* 413, 719–723 (2001).

11. Barrett, P. J. in *Glacial Processes and Products* (eds Hambrey, M. J. *et al.*). Vol. 39 259–287 (International Association of Sedimentologists, Special Publication, 2007).
12. Summerfield, M. A. *et al.* Cosmogenic isotope data support previous evidence of extremely low rates of denudation in the Dry Valleys region, southern Victoria Land. *Geol. Soc. Lond. Spec. Publ.* **162**, 255–267 (1999).
13. Lewis, A. R., Marchant, D. R., Ashworth, A. C., Hemming, S. R. & Machlus, M. L. Major middle Miocene global climate change: evidence from East Antarctica and the Transantarctic Mountains. *Geol. Soc. Am. Bull.* **119**, 1449–1461 (2007).
14. Marchant, D. R. *et al.* Formation of patterned ground and sublimation till over Miocene glacier ice, southern Victoria Land, Antarctica. *Geol. Soc. Am. Bull.* **114**, 718–730 (2002).
15. Drewry, D. J. *Antarctica: Glaciological and Geophysical Folio* (Scott Polar Research Institute, Univ. Cambridge, 1983).
16. Perkins, D. *Subglacial Landscape in Antarctica.* PhD thesis, Univ. Aberdeen (1984).
17. Watson, D. *Contouring: A Guide to the Analysis and Display of Spatial Data* 1–321 (Pergamon, 1992).
18. Tarboton, D. G., Bras, R. L. & Rodrigueziturbe, I. On the extraction of channel networks from digital elevation data. *Hydrol. Process.* **5**, 81–100 (1991).
19. Jamieson, S. S. R., Hulton, N. R. J. & Hagdorn, M. Modelling landscape evolution under ice sheets. *Geomorphology* **97**, 91–98 (2008).
20. Anderson, R. S., Molnar, P. & Kessler, M. A. Features of glacial valley profiles simply explained. *J. Geophys. Res.* **111**, F01004, DOI:10.1029/2005JF000344 (2006).
21. MacGregor, K. C., Anderson, R. S., Anderson, S. P. & Waddington, E. D. Numerical simulations of longitudinal profile evolution of glacial valleys. *Geology* **28**, 1031–1034 (2000).
22. Amerson, B. E., Montgomery, D. R. & Meyer, G. Relative size of fluvial and glaciated valleys in central Idaho. *Geomorphology* **93**, 537–547 (2008).
23. Montgomery, D. R. Valley formation by fluvial and glacial erosion. *Geology* **30**, 1047–1050 (2002).
24. Siegert, M. J., Taylor, J. & Payne, A. J. Spectral roughness of subglacial topography and implications for former ice-sheet dynamics in East Antarctica. *Glob. Planet. Change* **45**, 249–263 (2005).
25. Paterson, W. S. B. *The Physics of Glaciers* 3rd edn 1–480 (Pergamon, 1994).
26. Raine, J. I. & Askin, R. A. Terrestrial palynology of Cape Roberts drillhole CRP-3, Victoria Land Basin, Antarctica. *Terra Antarct.* **8**, 389–400 (2001).
27. Ehrmann, W. U., Setti, M. & Marinoni, L. Clay minerals in Cenozoic sediments off Cape Roberts (McMurdo Sound, Antarctica) reveal the palaeoclimatic history. *Palaeogeogr. Palaeoclim. Palaeoecol.* **229**, 187–211 (2005).
28. DeConto, R. M. & Pollard, D. A coupled climate-ice sheet modeling approach to the early Cenozoic histoy of the Antarctic ice sheet. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **198**, 39–53 (2003).
29. Harwood, D. M., McMinn, A. & Quilty, P. G. Diatom biostratigraphy and age of the Pliocene Sørsdal Formation, Vestfold Hills, East Antarctica. *Antarct. Sci.* **12**, 443–462 (2000).
30. Abramov, O. & McEwen, A. Technical note: An evaluation of interpolation methods for Mars Orbiter Laser Altimeter (MOLA) data. *Int. J. Remote Sens.* **25**, 669–676 (2004).

# LETTERS

# Kinematic variables and water transport control the formation and location of arc volcanoes

T. L. Grove[1], C. B. Till[1], E. Lev[1], N. Chatterjee[1] & E. Médard[1]†

The processes that give rise to arc magmas at convergent plate margins have long been a subject of scientific research and debate[1–6]. A consensus has developed that the mantle wedge overlying the subducting slab[3,4] and fluids and/or melts from the subducting slab itself[6–11] are involved in the melting process. However, the role of kinematic variables such as slab dip and convergence rate in the formation of arc magmas is still unclear. The depth to the top of the subducting slab beneath volcanic arcs, usually ~110 ± 20 km, was previously thought to be constant among arcs[3,6,12]. Recent studies[13,14] revealed that the depth of intermediate-depth earthquakes underneath volcanic arcs, presumably marking the slab–wedge interface, varies systematically between ~60 and 173 km and correlates with slab dip and convergence rate. Water-rich magmas (over 4–6 wt% $H_2O$) are found in subduction zones with very different subduction parameters, including those with a shallow-dipping slab (north Japan), or steeply dipping slab (Marianas). Here we propose a simple model to address how kinematic parameters of plate subduction relate to the location of mantle melting at subduction zones. We demonstrate that the location of arc volcanoes is controlled by a combination of conditions:

melting in the wedge is induced at the overlap of regions in the wedge that are hotter than the melting curve (solidus) of vapour-saturated peridotite and regions where hydrous minerals both in the wedge and in the subducting slab break down. These two limits for melt generation, when combined with the kinematic parameters of slab dip and convergence rate, provide independent constraints on the thermal structure of the wedge and accurately predict the location of mantle wedge melting and the position of arc volcanoes.

The extent and maximum depth of hydrous melting at subduction zones are controlled by the supply of water from the subducting slab to the overlying mantle wedge. New experimental work[15,16] shows that chlorite, $[Mg_{4.6}Fe_{0.3}Cr_{0.1}Al][Si_3Al]O_{10}(OH)_8$, which contains 13 wt% water, is stable in the mantle wedge to depths of 100–120 km. As water is released from the slab at shallow depths, it rises into overlying mantle and hydrates it through the formation of chlorite. This chlorite zone is advected to depth by corner flow as the sinking slab drags overlying mantle downward. This zone of chlorite stability in the wedge dehydrates at temperatures slightly above the vapour-saturated solidus. Water released from chlorite breakdown is then transferred
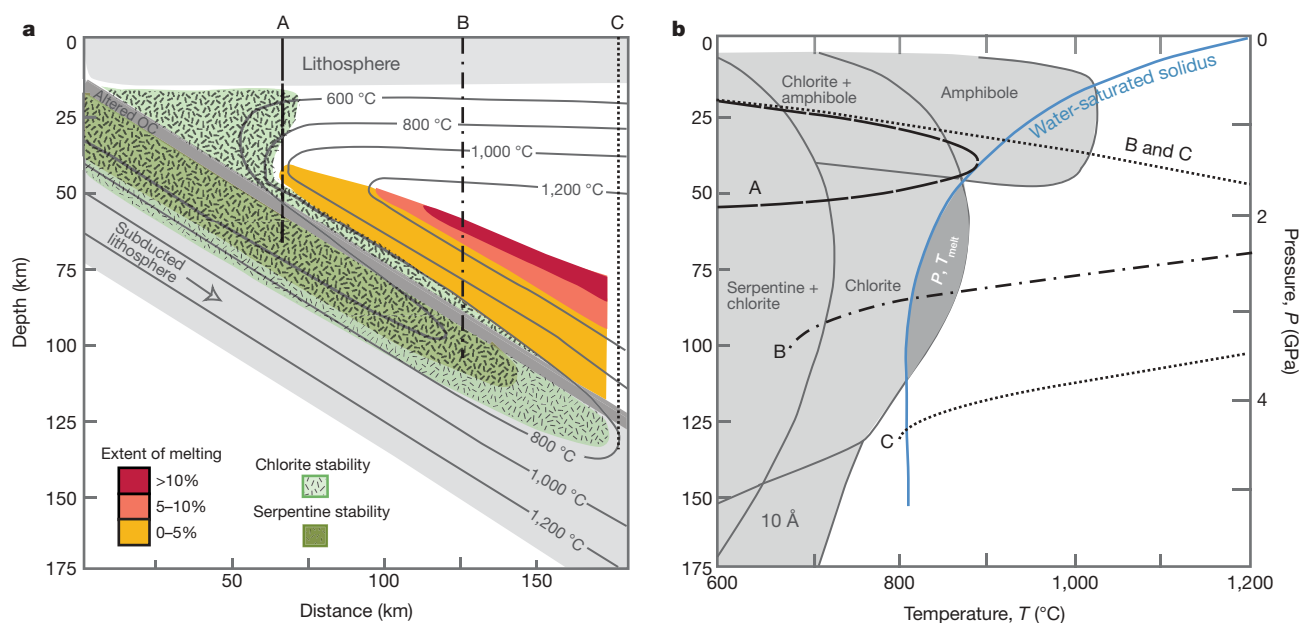


**Figure 1 | Diagrams showing limits for melt generation.** **a**, A typical subduction zone with melting model (after ref. 15) using our thermal model with a dip of 30° and a convergence rate of 40 mm yr⁻¹, and depicting the hydrous phase (>10 wt% water) stability fields within the subducted lithosphere and mantle wedge[15,16,18–22]. **b**, Hydrous peridotite phase diagram. In **a**, the vertical lines A, B and C denote vertical cross-sections through the mantle wedge and in **b** they denote a path through a pressure–temperature phase diagram for vapour-saturated peridotite. The dark-grey region in **b** labelled '$P$, $T_{melt}$' indicates the pressure–temperature region required for melting in our geodynamic models.

[1]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. †Present address: Laboratoire Magmas et Volcans, Université Blaise Pascal, Clermont-Ferrand, F-63038, France.

into a vapour-saturated melt (Fig. 1). The temperature of chlorite breakdown is ~200 °C hotter than temperature estimates of the slab–wedge interface for hot slab subduction[17,18]. Therefore, the presence of chlorite in the wedge provides a mechanism for transporting significant quantities of water (up to 2 wt% $H_2O$ in bulk peridotite) into the mantle[19] even in hot subduction zones and removes the need for dry melting in hot slab environments, as has been proposed for some subduction zones[9].

The maximum depth stability of hydrous minerals, and thus the maximum depth water released within the slab can ascend into the overlying mantle wedge, is influenced by the temperature structure in the subducted lithosphere (Fig. 1). Serpentine and the 10 Å phase are two hydrous minerals, in addition to chlorite, that are stable to considerable depths in the subducted oceanic lithosphere[18–21] (up to 150 km, depending on the pressure–temperature structure of the slab). Amphibole was once proposed as a carrier for transporting water to depth in the mantle[4,5], but amphibole becomes unstable at depths shallower than the depths at which melting begins beneath most arc volcanoes. As shown in the new vapour-saturated peridotite melting experiments[15,16,22], amphibole is stable from ~0.8 to 2 GPa on the solidus and chlorite is stable from 2 to 3.6 GPa (ref. 16; Fig. 1b).

Here we incorporate these new observations regarding the stability of hydrous minerals and the vapour-saturated peridotite solidus to estimate the effect of kinematic subduction parameters on the distribution of flux melting in subduction zones. Many studies indicate that convergence rate and the dip of the downgoing slab at subduction zones probably strongly affect the location of arc volcanoes at convergent margins[13,14,23–25]. The convergence rate influences the advection of heat by corner flow in the mantle wedge and the diffusion of heat across the steep temperature gradient induced between the cold slab and the hot upper mantle[5,13,26–29]. Slab dip determines the geometry of the flow of hot mantle into the wedge corner. To isolate and constrain the effects of slab dip and convergence rate on the location and size of

the melting region at convergent margins, we construct a set of geodynamic models.

We construct two-dimensional finite-element models of kinematically driven subduction following the set-up defined in the subduction community benchmark model[30]. Slab dips in our model span the range found for most subduction zones and vary between 32° and 64°. In the same way, we vary slab velocities from 20 to 100 mm $yr^{-1}$. For each combination of parameters, we calculate the steady-state thermal structure of the slab and wedge. Mantle rheology in our models is Newtonian, isotropic and temperature-dependent, as is appropriate for the upper mantle. Grid resolution is 2.3 km in both the vertical and horizontal directions.

After finding the thermal structure of the slab and the mantle wedge, we calculate the region of potential melting by comparing the stability regions of chlorite and the vapour-saturated solidus of peridotite[15,16]. When water ascends by porous flow into the hotter, shallower parts of the mantle wedge, melting begins at a critical distance from the trench where temperature first exceeds that of the vapour-saturated solidus (Fig. 1, path A). Fluid ascent paths closer to the wedge corner never reach temperatures that exceed the vapour-saturated solidus (Fig. 1, to the left of path A). Between paths A and C in Fig. 1, melting may occur by one of two mechanisms (Fig. 1, path B). If a water-rich fluid is present, melting will begin at the vapour-saturated solidus. If there is no free fluid phase, chlorite breakdown, at a temperature slightly above the vapour-saturated solidus, will initiate melting. In our models, we have chosen the region of melting to be above the vapour-saturated solidus but below the chlorite breakdown reaction between 2 and 3.6 GPa (~60 and 110 km depth, respectively) (region labelled '$P$, $T_{melt}$' in Fig. 1b). Above 3.6 GPa (~110 km), little to no melting will occur because the water necessary for vapour-saturated melting is not present (Fig. 1, path C) as the upper-temperature limit for hydrous phase stability plunges below the vapour-saturated solidus. (At pressures below 2 GPa, vapour-saturated melting might be possible if
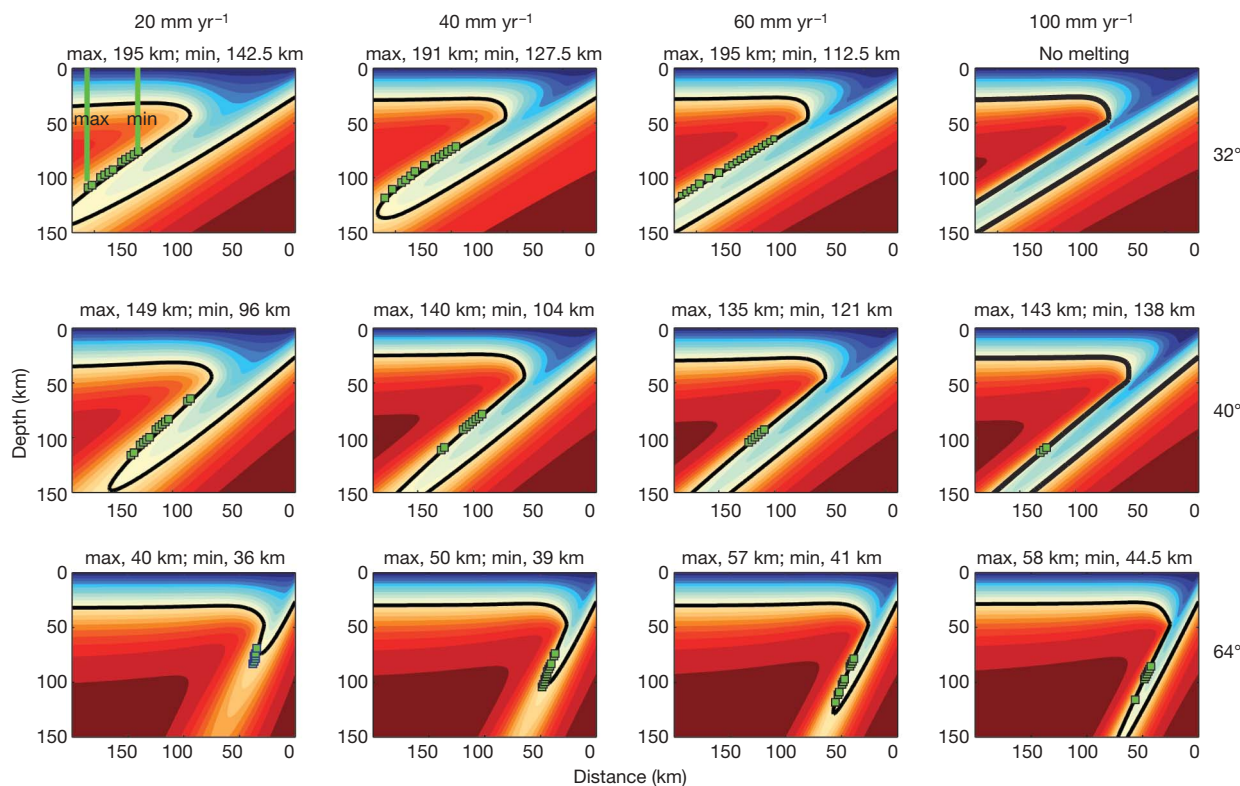


**Figure 2 | Thermal structure of the mantle wedge and subducting slabs calculated from finite-element modelling and the location of the initiation of vapour-saturated melting.** Rows have constant slab dip (in degrees) and columns have constant slab velocity (convergence rate in mm $yr^{-1}$). Green squares denote the initiation of hydrous melting by marking elements above

the temperature of the vapour-saturated peridotite solidus and within the chlorite stability field (region labelled '$P$, $T_{melt}$' in Fig. 2). The thick black contour follows the 800 °C isotherm. The values above each plot state the minimum and maximum distances (in kilometres) of melting from the trench (defined in top left plot).

**Table 1 | Outputs from our geodynamic models with measured arc volcano locations**
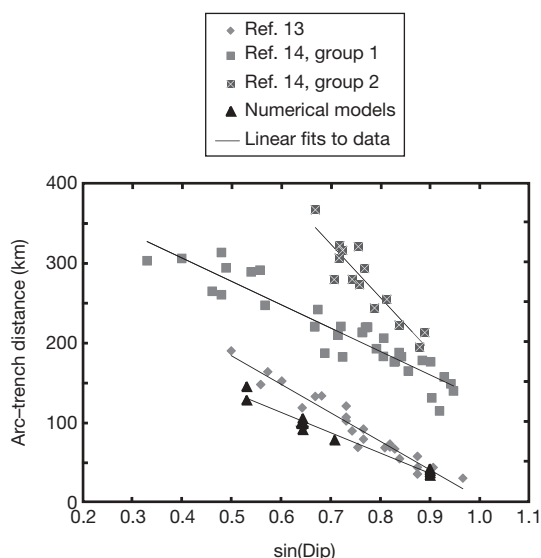
| Arc name | Convergence rate (km per million years) | Slab dip (degrees) | Depth to slab (km) | Arc–trench distance (km) |
|---|---|---|---|---|
| Central Aleutians | 60 | 58 | 183 (ref. 14); 44 (ref. 13) | 90 (ref. 14); 90 (ref. 13) |
| Model similar to the central Aleutians | 60 | 64 | 40 | 80 |
| New Zealand | 36 | 50 | 96 (ref. 14); 92 (ref. 13) | 123 (ref. 14); 110 (ref. 13) |
| Model similar to New Zealand | 40 | 40 | 104 | 70 |

Data for modern subduction zones from ref. 13 did not contain direct measurements of distance from trench to arc. We calculate this value from their measurements of depth to slab and the slab dip by a simple geometric approximation: (depth to slab)/tan (slab dip).

the mantle wedge temperature exceeds the vapour-saturated solidus. We have chosen not to examine this process in our model.)

The thermal structures resulting from the range of slab dips and convergence rates we examine are depicted in Fig. 2. Our modelling results suggest the dip of the slab is an order of magnitude more important than the convergence rate in controlling the location of the isotherms in the mantle at subduction zones. As the slab steepens, hotter mantle can reach shallower depths in the mantle wedge corner (Fig. 2), leading to larger extents of melting. Additionally, slab dip has a first-order effect on the total width of the melting region: shallower slabs lead to wider melting regions, suggesting a correlation between slab dip and arc width. Jarrard[12] and Tatsumi[6] noted a similar relationship between slab dip and arc width. Tatsumi attributed it to the dehydration of hydrous minerals in the subducting lithosphere. Our model and experiments indicate that this correlation results from chlorite dehydration and melting in the mantle wedge.

Our models reveal a clear relationship between slab dip and the position of the melting region, a relationship we confirm by observations from subduction zones worldwide. For steeply dipping slabs, the melting region is closer to the trench and the surface width of the melting region (that is, the width of the volcanic arc) is narrower than for shallow-dipping slabs (Fig. 2). A plot of the distance from the trench to the volcanic arc front (which we assume to lie directly above the edge of the melting region) as a function of the sine of the slab dip reveals a clear linear trend with a negative correlation. Observations from arcs worldwide, collected by England *et al.*[13] and Syracuse and Abers[14] confirm this linear negative correlation between trench–arc distance and the sine of slab dip that we find in our models (Fig. 3). The slope

from our models is similar to the slopes obtained from refs 13 and 14. This agreement demonstrates a fundamental relationship between slab dip and the location of melting and arc volcanism. Interestingly, part of the data set of ref. 14 ('group 2') shows a secondary steeper linear trend, which has yet to be explained.

The correlation between convergence rate and the location of the melting region or melt productivity is less clear. In general, there is no systematic influence of convergence rate on the width of the melting region (Fig. 2). At a dip of 32°, the melting region systematically widens with increasing convergence rate, whereas at a dip of 40°, the melting region narrows.

Predictions of the location of melting from our geodynamic models for subduction zones correspond well with the locations of the central Aleutian and New Zealand arcs. In Table 1, the locations of volcanic front for the central Aleutian and New Zealand arcs are compared to geodynamic models we constructed with similar kinematic variables. In both cases, our models predict results within 10 km of the observed locations of arc volcanoes, suggesting our model is a good first-order approximation of the effect of kinematic variables on the thermal structure of the mantle wedge and the location of melting at subduction zones.

In conclusion, the petrologic control on mantle melting at subduction zones is the temperature of the vapour-saturated peridotite solidus and the supply of water to the mantle wedge from the breakdown hydrous minerals, predominantly chlorite in the mantle wedge and subducted lithosphere. The kinematic control on the location of mantle melting is primarily slab dip and, to a lesser degree, the convergence rate. Thus, regions of the mantle wedge saturated with fluid and above the temperature of the vapour-saturated peridotite solidus will melt, and the size and location of this melting region will vary linearly with slab dip. Our experimental work and models adequately explain worldwide observations regarding the location of arc volcanoes and clarify the effect of the kinematic variables, such as slab dip and convergence rate, on the formation of arc magmas.

**Figure 3 | Distance between the arc front and trench versus the sine of slab dip for subduction zones worldwide and for our numerical models.** Models (shown as triangles) follow the same trend as the data of ref. 13 and those from South and Central America in ref. 14 ('group 1'). Data from Alaska and the western Pacific in ref. 14 ('group 2') follow a second, steeper trend. The constant offset of ~100 km between the ref. 14 'group 1' data and the data from ref. 13 are explained by the different techniques used to estimate the trench position. Solid lines are linear fits to the data.

1. Toksoz, M., Minear, J. & Julian, B. Temperature field and geophysical effects of a downgoing slab. *J. Geophys. Res.* **76**, 1113–1138 (1971).
2. Ringwood, A. The petrological evolution of island arc systems. *J. Geol. Soc.* **130**, 183–204 (1975).
3. Gill, J. B. *Orogenic Andesites and Plate Tectonics* 390 (Springer, New York, 1981).
4. Wyllie, P. Subduction products according to experimental prediction. *Geol. Soc. Am.* **93**, 468–476 (1982).
5. Davies, J. H. & Stevenson, D. J. Physical model of source region of subduction zone volcanics. *J. Geophys. Res.* **97**, 2037–2070 (1992).
6. Tatsumi, Y. The subduction factory: how it operates in the evolving Earth. *GSA Today* **15**, 4–10 (2005).
7. Moran, A. E., Sisson, V. B. & Leeman, W. P. Boron depletion during progressive metamorphism—implications for subduction processes. *Earth Planet. Sci. Lett.* **111**, 331–349 (1992).
8. Yogodzinski, G. M., Kay, R. W., Volynets, O. N., Koloskov, A. V. & Kay, S. M. Magnesian andesite in the western Aleutian Komandorsky region - implications for slab melting and processes in the mantle wedge. *Geol. Soc. Am.* **107**, 505–519 (1995).
9. Stern, C. R. & Kilian, R. Role of the subducted slab, mantle wedge and continental crust in the generation of adakites from the Andean Austral volcanic zone. *Contrib. Mineral. Petrol.* **123**, 263–281 (1996).
10. Kelemen, P. B., Hart, S. R. & Bernstein, S. Silica enrichment in the continental upper mantle via melt/rock reaction. *Earth Planet. Sci. Lett.* **164**, 387–406 (1998).
11. Plank, T. & Langmuir, C. H. Tracing trace-elements from sediment input to volcanic output at subduction zones. *Nature* **362**, 739–743 (1993).

12. Jarrard, R. D. Relations among subduction parameters. *Rev. Geophys.* **24,** 217–284 (1986).
13. England, P., Engdahl, R. & Thatcher, W. Systematic variation in the depths of slabs beneath arc volcanoes. *Geophys. J. Int.* **156,** 377–408 (2004).
14. Syracuse, E. M. & Abers, G. A. Global compilation of variations in slab depth beneath arc volcanoes and implications. *Geochem. Geophys. Geosyst.* **7,** Q05017, doi:10.1029/2005GC001045 (2006).
15. Grove, T. L., Chatterjee, N., Parman, S. W. & Medard, E. The influence of $H_2O$ on mantle wedge melting. *Earth Planet. Sci. Lett.* **249,** 74–89 (2006).
16. Till, C. B., Grove, T. L., Withers, A. C. & Hirschmann, M. M. Extending the wet mantle solidus: implications for $H_2O$ transport and subduction zone melting processes. *AGU Fall Meet. Suppl.* Abstract D142A–02 (2007).
17. Kelemen, P. B., Rilling, J. L., Parmentier, E. M., Mehl, L. & Hacker, B. R. in *Inside the Subduction Factory* (ed. Eiler, J. M.) 293–311 (American Geophysical Union, 2003).
18. Pawley, A. Chlorite stability in mantle peridotite: the reaction clinochlore plus enstatite=forsterite+pyrope+$H_2O$. *Contrib. Mineral. Petrol.* **144,** 449–456 (2003).
19. Iwamori, H. Transportation of $H_2O$ and melting in subduction zones. *Earth Planet. Sci. Lett.* **160,** 65–80 (1998).
20. Schmidt, M. W. & Poli, S. Experimentally based water budgets for dehydrating slabs and consequences for arc magma generation. *Earth Planet. Sci. Lett.* **163,** 361–379 (1998).
21. Fumagalli, P. & Poli, S. Experimentally determined phase relations in hydrous peridotites to 6.5 GPa and their consequences on the dynamics of subduction zones. *J. Petrol.* **46,** 555–578 (2005).
22. Medard, E. & Grove, T. L. Early hydrous melting and degassing of the Martian interior. *J. Geophys. Res.* **111,** E1103 (2006).
23. Kincaid, C. & Sacks, I. S. Thermal and dynamical evolution of the upper mantle in subduction zones. *J. Geophys. Res.* **102,** 12295–12315 (1997).
24. van Keken, P. E., Kiefer, B. & Peacock, S. M. High-resolution models of subduction zones: implications for mineral dehydration reactions and the transport of water into the deep mantle. *Geochem. Geophys. Geosyst.* **3,** doi:10.1029/2001GC000256 (2002).
25. England, P. & Wilkins, C. A simple analytical approximation to temperature structure in subduction zones. *Geophys. J. Int.* **159,** 1138–1154 (2004).
26. Peacock, S. M., Rushmer, T. & Thompson, A. B. Partial melting of subducting oceanic-crust. *Earth Planet. Sci. Lett.* **121,** 227–244 (1994).
27. Kirby, S. H., Durham, W. B. & Stern, L. A. Mantle phase-changes and deep-earthquake faulting in subducting lithosphere. *Science* **252,** 216–225 (1991).
28. Molnar, P. & England, P. Temperatures, heat-flux, and frictional stress near major thrust faults. *J. Geophys. Res.* **95,** 4833–4856 (1990).
29. McKenzie, D. Speculations on the consequences and causes of plate motions. *Geophys. J. R. Astron. Soc.* **18,** 1–32 (1969).
30. van Keken, P. E. *et al.* A community benchmark for subduction zone modeling. *Phys. Earth Planet. Inter.* **171,** 187–197 (2008).

# LETTERS

# Parvalbumin neurons and gamma rhythms enhance cortical circuit performance

Vikaas S. Sohal[1]*, Feng Zhang[1]*, Ofer Yizhar[1] & Karl Deisseroth[1]

Synchronized oscillations and inhibitory interneurons have important and interconnected roles within cortical microcircuits. In particular, interneurons defined by the fast-spiking phenotype and expression of the calcium-binding protein parvalbumin[1,2] have been suggested to be involved in gamma (30–80 Hz) oscillations[3–7], which are hypothesized to enhance information processing[8,9]. However, because parvalbumin interneurons cannot be selectively controlled, definitive tests of their functional significance in gamma oscillations, and quantitative assessment of the impact of parvalbumin interneurons and gamma oscillations on cortical circuits, have been lacking despite potentially enormous significance (for example, abnormalities in parvalbumin interneurons may underlie altered gamma-frequency synchronization and cognition in schizophrenia[10] and autism[11]). Here we use a panel of optogenetic technologies[12–14] in mice to selectively modulate multiple distinct circuit elements in neocortex, alone or in combination. We find that inhibiting parvalbumin interneurons suppresses gamma oscillations *in vivo*, whereas driving these interneurons (even by means of non-rhythmic principal cell activity) is sufficient to generate emergent gamma-frequency rhythmicity. Moreover, gamma-frequency modulation of excitatory input in turn was found to enhance signal transmission in neocortex by reducing circuit noise and amplifying circuit signals, including inputs to parvalbumin interneurons. As demonstrated here, optogenetics opens the door to a new kind of informational analysis of brain function, permitting quantitative delineation of the functional significance of individual elements in the emergent operation and function of intact neural circuitry.

We first developed a versatile system to selectively express microbial opsins, enhanced *Natronomonas pharaonis* halorhodopsin (eNpHR) or channelrhodopsin-2 (ChR2), in fast-spiking parvalbumin (PV) interneurons. Although opsin expression can be targeted with a PV promoter fragment[15], the resulting level of expression was insufficient to drive action potentials reliably (data not shown). We therefore devised (Supplementary Information) a Cre-recombinase-dependent adeno-associated virus (AAV) expression system carrying a reversed[16] and doubled-floxed transgene to enable selective expression in transgenic mice expressing Cre recombinase under the control of the PV promoter (*PV::Cre*). This approach decouples transcriptional strength from the specificity of the PV promoter, allowing expression of microbial opsins driven by the strong *EF1α* promoter (Supplementary Information). We found leak to be undetectable and the opsin expression level to be robust (Supplementary Fig. 1a–c), indicating the quantitative suitability of this strategy for probing the causal role of these neurons in neocortical processing.

Cre-dependent *eNpHR–eYFP* (enhanced yellow fluorescent protein) or *ChR2–eYFP* AAV5 vectors were stereotactically injected into the cortex of *PV::Cre* transgenic mice. Numerous interneurons with

eYFP fluorescence were observed (Fig. 1c, left); in contrast, injection of this construct into wild-type mice revealed no detectable expression (Fig. 1c, right panels), indicating specificity to PV interneurons. Although PV co-staining is less reliable (PV expression in fast-spiking cells can drop below the staining detection limit[17]), immunohistochemistry also revealed both co-localization of eYFP-tagged microbial opsins with PV immunoreactive cells and successful transduction of a substantial fraction of PV cells (Supplementary Fig. 1d; note that the PV interneuron population is heterogenous and not all PV interneurons necessarily express Cre in a given line). As anticipated from the robust expression achieved with the double-floxed strategy, cells expressing either eNpHR–eYFP or ChR2–eYFP displayed strong photocurrents (Supplementary Fig. 1e, f). In particular, in PV interneurons expressing eNpHR–eYFP, yellow light elicited hyperpolarizing currents that inhibited spiking (Fig. 1d, e).

We used this system to test whether PV interneuron activity could be involved in gamma oscillation generation *in vivo*. We injected *Camk2α* (*CaMKIIα*)::*ChR2–eYFP* and Cre-dependent *eNpHR–eYFP* AAV5 into the prefrontal cortex of *PV::Cre* transgenic mice, yielding simultaneous ChR2 expression in pyramidal (PY) neurons and eNpHR expression in fast-spiking PV interneurons (Fig. 1f). Local field potentials *in vivo* (Fig. 1g) revealed that single flashes of blue light delivered in naturalistic trains were followed by gamma oscillations (Fig. 1h–j) that were strongly phase-locked to flashes (Fig. 1k); evoked activity was largely limited to the gamma band. We used wavelet methods (Supplementary Information) to measure power at various frequencies after blue flashes, and found that yellow light suppressed gamma power (Fig. 1h, l; $20 \pm 4\%$, mean $\pm$ s.e.m.), but did not reduce non-gamma power ($P < 0.001$; recordings from 4 different locations in 2 mice; inhibition is in the range expected from the observations that ~41% of PV interneurons express eNpHR under these conditions and that yellow light suppresses ~65% of PV spikes).

Because inhibition of PV interneurons was found to suppress gamma power, we next sought to determine whether stimulating PV cells could elicit gamma oscillations in downstream PY neurons. To achieve precise and specific control of inputs to PV and PY neurons, we used brain slices from *PV::Cre* mice injected with Cre-dependent *ChR2–eYFP* AAV5 in prefrontal cortex. Blue light drove spikes in PV interneurons and inhibited spikes in PY cells, as expected (Fig. 2a, b), in a GABA_A (γ-aminobutyric acid subtype A)-receptor-dependent manner (blocked by 50 μM picrotoxin; data not shown). Next we used dynamic-clamp technique to inject well-defined trains of simulated excitatory post-synaptic currents (sEPSCs) into PY neurons, and drove fast-spiking PV interneurons by delivering light flashes triggered by observed PY neuron spikes, thereby implementing classical feedback inhibition (Fig. 2c; the delay between spike detection and onset of inhibitory post-synaptic responses was

[1]Department of Bioengineering, Department of Psychiatry and Behavioral Sciences, W083 Clark Center, 318 Campus Drive West, Stanford University, Stanford, California 94305, USA.
*These authors contributed equally to this work.
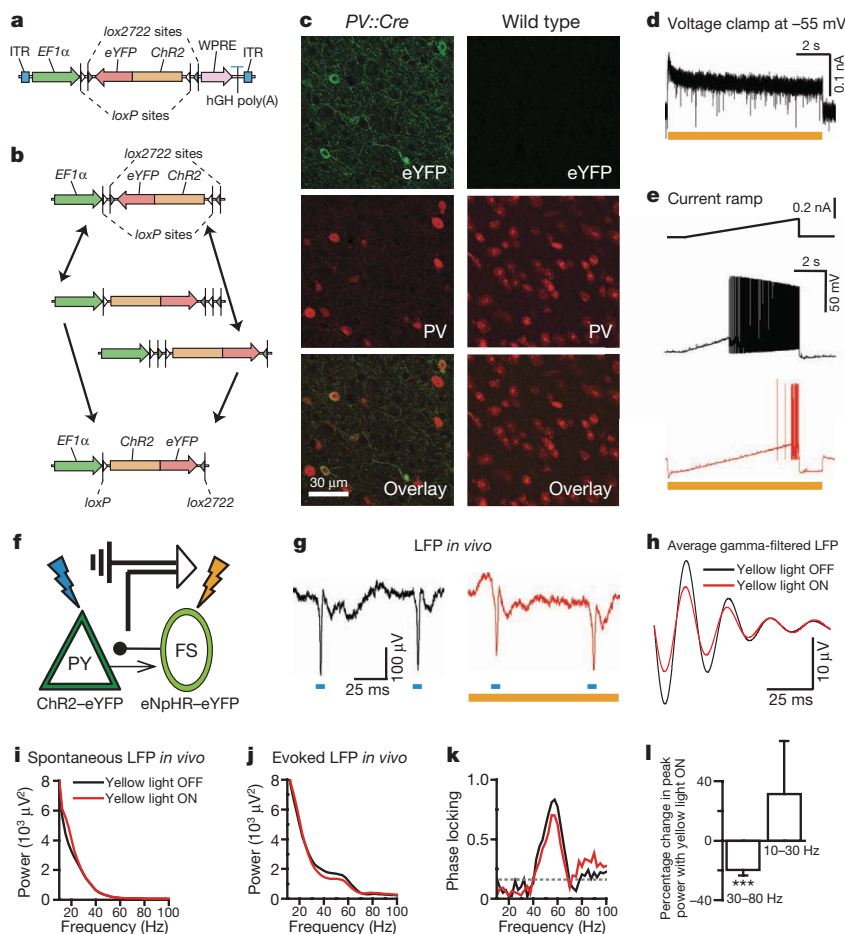
**Figure 1 | Inhibiting PV cells suppresses gamma oscillations *in vivo*.**
**a**, Double-floxed Cre-dependent AAV vector design. hGH, human growth
hormone polyadenylation signal; ITR, inverted terminal repeat; WPRE,
woodchuck heptatitis virus post-transcriptional regulatory element.
**b**, Recombination pathways to low-leak Cre-dependent expression. **c**, Left:
antibody-stained PV cells (red) expressing ChR2–eYFP (green) in prefrontal
cortex (PFC) of *PV::Cre* mice after injection of Cre-dependent AAV. Right:
absent ChR2–eYFP expression in similarly treated wild-type mice. **d**, Yellow-
light-evoked outward current in eNpHR-expressing PV cells in acute slice.
**e**, Effect of yellow light on current-ramp-evoked spiking in the same cell.

**f**, Experimental design: *in vivo* local field potentials (LFPs) recorded in
mouse PFC; blue and yellow light modulate ChR2(+) PY and eNpHR(+)
fast-spiking (FS)/PV cells, respectively. **g**, Sample blue-light-evoked LFPs;
**g**–**k** show data from a single location *in vivo*; red traces denote recordings in
yellow light. **h**, Filtered (35–45 Hz)-light-evoked LFPs. **i**, Spontaneous LFP
power spectra. **j**, Blue-light-evoked LFP power spectra. **k**, Yellow light
modulation of phase-locking (Supplementary Information). Dotted line
denotes statistical significance. **l**, Effect of fast-spiking cell inhibition on peak
power at gamma and lower frequencies ($n = 4$ recording locations;
***$P < 0.001$). Error bars, mean and s.e.m.

~3 ms and light is predicted to drive ~2–4 PV interneurons per field;
Supplementary Information). Notably, during responses to entirely
non-rhythmic sEPSC trains (Fig. 2d), PV interneuron-mediated
feedback inhibition by itself markedly enhanced power in the gamma
range, but not at other frequencies (representative cell in Fig. 2e;
population data in Fig. 2f; $n = 4$ cells, $P < 0.05$ by two-tailed *t*-test
here and in other figures). Quantitatively comparable levels of feed-
back inhibition delivered with dynamic clamp to PY cells rather than
with PV interneuron stimulation were not effective at enhancing
gamma power ($n = 4$ cells; Supplementary Tables 1 and 2), pointing
to a specialized role of PV interneurons in generating emergent
gamma oscillations within neocortical microcircuits.

We next sought to quantify effects of gamma oscillations on
information processing. We used dynamic-clamp technique to drive
PY neurons ($n = 14$) with precisely defined trains of sEPSCs (Fig. 3a).
During each train, the rate of input sEPSCs was varied from 0 Hz to
500 Hz; to create rhythmic inputs, each sEPSC train was further
modulated by means of 25-ms cycles (40 Hz; gamma) or 125-ms cycles
(8 Hz; theta)[18,19]. For each condition, we measured output spike rate as
a function of input sEPSC rate (representative neuron in Fig. 3b). In all
conditions, mean spike rate increased with sEPSC rate, and gamma
(but not theta) oscillations increased the gain—that is, the maximal
slope—of the neuronal input–output curve (Fig. 3c; $P < 0.001$, $n = 14$

cells). Both gamma and theta modulation also significantly reduced
response variability (mean normalized standard deviation of res-
ponses to each input sEPSC rate; Supplementary Information;
Fig. 3d; $P < 0.001$). Detailed quantification of the effect of rhythmic
modulation requires informational measures, because it is otherwise
not clear whether (for example) steepening of the input–output curve
for input rates near 200 Hz, or flattening of the curve near 400 Hz, has
a greater impact on information transmission properties.

Mutual information[20] can quantify the information a neural output
carries about an input signal, defined as the difference between res-
ponse entropy (variability of output) and noise entropy (how much
output variability is unrelated to input; Supplementary Information).
Both theta and gamma modulation markedly increased mutual
information between input sEPSC rate and output spike rate
(Fig. 3e). In particular, gamma modulation increased the amount of
information that spike rates transmitted about the number of sEPSCs
over gamma-cycle time windows (non-rhythmic, $0.43 \pm 0.05$ bits;
gamma, $0.62 \pm 0.04$ bits; $P < 0.001$) and on longer timescales
(theta-cycle time window: non-rhythmic, $1.17 \pm 0.07$ bits; gamma,
$1.37 \pm 0.04$ bits; $P < 0.001$). In contrast, theta modulation increased
sEPSC spike rate information only on theta-cycle timescales and
longer (Fig. 2e). Similar results were obtained from both frontal
(above) and prefrontal ($n = 5$ cells) layer V/VI pyramidal neurons,
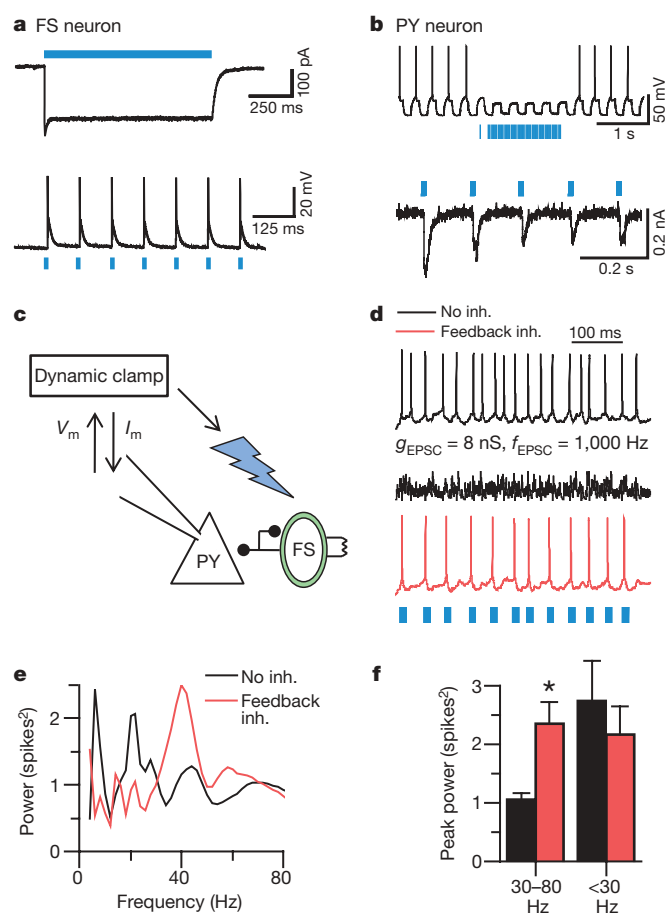
**Figure 2 | Feedback inhibition from PV cells generates emergent gamma frequency synchrony. a**, Light-evoked responses in a fast-spiking PV interneuron (FS). **b**, Responses of a ChR2(−) PY cell during photoactivation of ChR2(+) PV interneurons. **c**, Experimental design: sEPSCs drive PY cells. Light flashes triggered by PY cell spikes activate FS/PV interneurons (optical feedback inhibition). $V_m$, measured membrane potential; $I_m$, injected current. **d**, PY cell responses to non-rhythmic sEPSCs with and without this optical feedback inhibition (inh.). $g_{EPSC}$, unitary sEPSC conductance; $f_{EPSC}$, sEPSC frequency. **e**, Power spectra obtained by convolving the spike trains of a PY cell with wavelets of varying frequencies; red trace represents optical feedback inhibition via PV interneurons. **f**, Summary of spectral data at gamma (30–80 Hz) and lower frequencies ($n = 4$ cells; *$P < 0.05$). Error bars, mean and s.e.m.



**Figure 3 | Gamma oscillations amplify signals and reduce noise in PY cells. a**, PY cell responses (top traces) to non-rhythmic (NR), repeated non-rhythmic (Rep), or rhythmic defined sEPSC trains (lower traces; Supplementary Information; depicted rates = 40–372 Hz). **b**, Spike rates of representative PY cell under each rhythmicity condition. **c**, Maximum input–output (I–O) gain. **d**, Response variability for each condition. **e**, Left: mutual information between output spike number and input sEPSCs; gamma oscillations consistently enhanced mutual information. Right: contrasting effect on sEPSC spike rate information in simulated integrate-and-fire cells using the same sEPSC trains ($n = 14$ cells in **c–e**; ***$P < 0.001$). Key in **e** is the same as for **b**. Error bars, mean and s.e.m.

and findings were robust to altering both sEPSC amplitudes and extents of oscillatory modulation (Supplementary Fig. 2). Information transmission changes could be accounted for entirely by decreased noise entropy (Supplementary Information; no increase in response entropy was observed), and this informational response to gamma-modulated input was not simply an effect of temporal summation, because simulated integrate-and-fire neurons receiving identical sEPSC trains did not enhance sEPSC spike rate information in response to gamma oscillations (Fig. 3e); instead, informational effects probably depend on specialized active conductances influencing native PY cells, as illustrated by biophysical models (Supplementary Fig. 8).

To extend these findings to trains of real EPSCs in large networks of neurons, we stimulated distributed populations of PY neurons in brain slices from transgenic mice expressing ChR2 under the *Thy1* promoter[21–23]. Specific expression in layer V pyramidal neurons and their processes (Fig. 4a), absence of light-evoked responses outside of layer V and absence of short-latency axonal-type responses to light all confirm that ChR2 activation was specific to layer V PY neurons, which comprise an interconnected network[24] that may maintain prefrontal

activity during working memory[25] and generate UP states of persistent network activity[26]. We recorded from multiple layer V cell types, and classified neurons on the basis of morphology, intrinsic membrane properties and light responses; in particular, we used the fast-spiking phenotype to identify PV interneurons[1,2], which did not express eYFP (Fig. 4a) and were not directly activated by light (Fig. 4c, e; Supplementary Information). We delivered rhythmic and non-rhythmic trains of stimulation to the spatially distributed PY circuit population by full-field light flashes via a ×40 objective rather than by dynamic clamp. This approach recruits a small distributed population of neurons (~20–50 PY neurons, Supplementary Information) in a manner that has been useful for studying temporal aspects of circuit function in other circuits such as the retina[27].

Modulating trains of light flashes using theta or gamma frequency patterns induced pronounced rhythmic firing at the corresponding frequencies (Fig. 4f) in PY neurons ($n = 12$), as expected. A valuable property of these virtually identical mice from the transgenic ChR2 line is that spike trains are remarkably similar in different pyramidal neurons from different slices (Fig. 4d). We took advantage of this fact (Supplementary Information) to compute the information that fast-spiking interneuron responses convey about the output of PY neurons. Both theta and gamma oscillations increased the amount of information transmitted from the PY neuron population to post-synaptic fast-spiking interneurons (Fig. 4g, h; gamma, $P < 0.001$; theta, $P < 0.05$; $n = 7$ cells), and once again we found that gamma oscillations enhanced information transmission across timescales (for example, over 125 ms: non-rhythmic, $1.02 \pm 0.11$ bits; gamma, $1.44 \pm 0.15$ bits; $P < 0.001$, $n = 7$ cells). We probed regular-spiking interneurons as well (regular-spiking here refers to non-fast-spiking
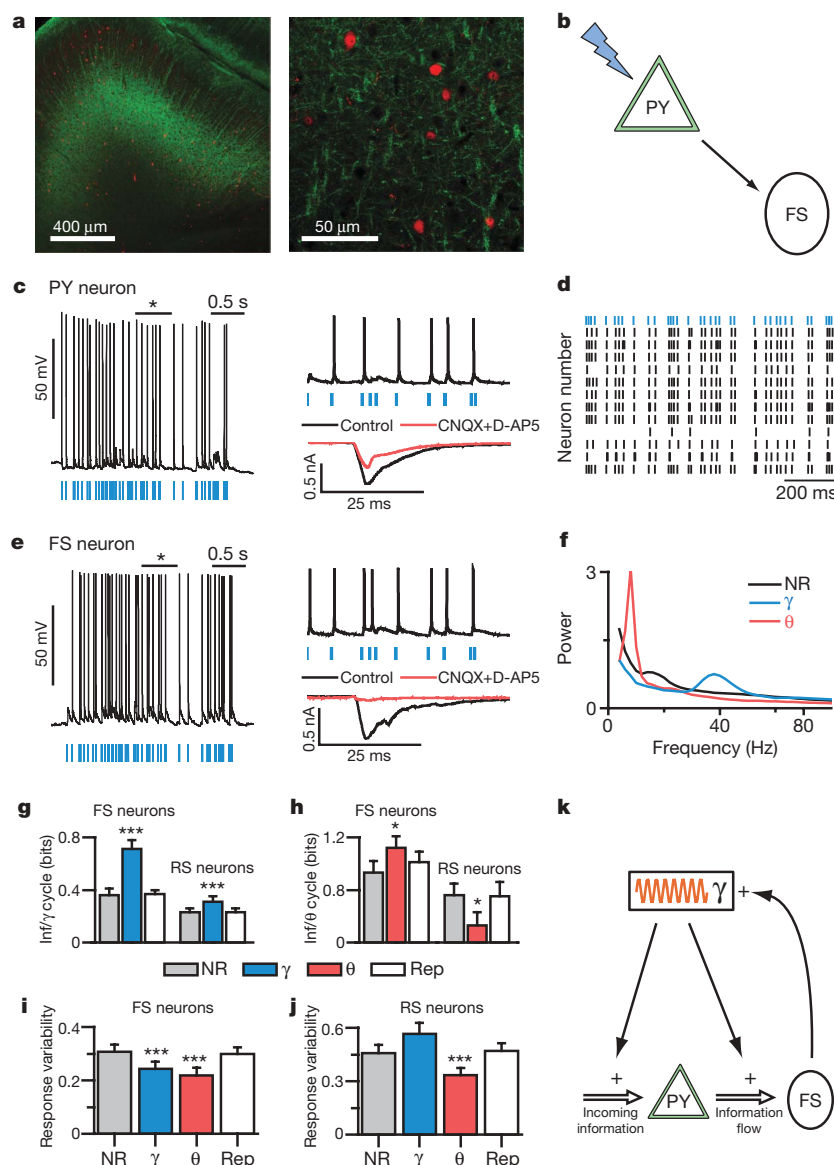
**Figure 4 | Gamma oscillations enhance information flow from PY to PV cells. a**, eYFP (green) and PV (red) cells in layer V PFC of *Thy1−ChR2–eYFP* transgenic mice. **b**, Experimental design: light flashes excite PY neurons, which synaptically excite PV cells. **c**, Light directly excites PY cells. Current evoked by 1-ms flashes with/without synaptic blockers (CNQX and D-AP5). **d**, Responses of PY neurons from different slices to the same light train (blue). **e**, Light indirectly excites fast-spiking (FS) interneurons. Current evoked by 1-ms flashes with/without synaptic blockers. **f**, Power spectra of PY neuron spiking elicited by distinct-rhythmicity light trains. **g, h**, Mutual information (Inf) between responses of fast-spiking or regular-spiking (RS) neurons and PY neurons to the same rhythmic or non-rhythmic light trains (Supplementary Information). **i, j**, Response variability of fast-spiking and regular-spiking cells for rhythmic and non-rhythmic stimuli ($n = 12$ PY, 7 fast-spiking, and 9 regular-spiking cells). **k**, Interplay of gamma ($\gamma$) oscillations and fast-spiking neurons. Error bars, mean and s.e.m.

neurons not directly activated by light; $n = 9$), and found surprising specificity of information effects to fast-spiking neurons; specifically, theta oscillations actually decreased information transmission through regular-spiking neurons (Fig. 4g, h, $P < 0.001$) and gamma oscillations exerted little effect on longer timescales (125 ms; $P = 0.82$, $n = 9$ cells). As in dynamic-clamp experiments (Fig. 3), oscillations reduced response variability (Fig. 4i, j, $P < 0.001$, $n = 7$ cells) and increased mutual information by decreasing noise entropy (noise entropies for 125-ms window: non-rhythmic, $1.73 \pm 0.22$ bits; gamma, $1.40 \pm 0.26$ bits; theta, $1.42 \pm 0.22$ bits; $P < 0.01$ and $P < 0.001$ for gamma and theta modulation, respectively; no increase in response entropy was observed). Decreases in noise entropy may relate to greater stereotypy in interspike intervals (ISIs) in PY neurons during rhythmic versus non-rhythmic states (Supplementary Fig. 3c). Indeed, by making ISIs more stereotyped, rhythmic modulation of inputs may reduce variability resulting from short-term

plasticity and feedback inhibition, which are likely relevant to the effects of oscillations in neural circuits.

Pioneering work has suggested a role for fast-spiking PV interneurons in important aspects of information processing in the brain[7]. However, it has been unclear whether these interneurons are necessary or sufficient to promote gamma oscillations, and what causal role these oscillations would have in modulating information processing. Although further work will be needed to differentiate among different subtypes of PV interneurons—for example, basket and axo-axonic cells, which may have different functions[28]—we observed and quantified significant functional synergy between fast-spiking PV interneurons and gamma oscillations (Fig. 4k) in promoting transmission of signals within neocortical microcircuits. Through these mechanisms, gamma oscillations generated by fast-spiking PV interneurons could amplify signals corresponding to attended stimuli[29] and facilitate the transfer of information both within and between brain regions[30], while

corresponding abnormalities in this process may contribute to the disabling symptoms of schizophrenia[10] and autism[11]. The triad of optogenetics, dynamic clamp and informational analysis enables precise control over input data integrated with precise tracking and measurement of relevant output data. Simultaneous control over these input and output data streams may enable a novel kind of reverse engineering of microcircuits in both health and disease, by facilitating quantitative insight into how interacting elements give rise to normal and pathological circuit function.

## METHODS SUMMARY

**Viral-mediated opsin expression.** We used a double-floxed inverted open reading frame strategy (Fig. 1a and Supplementary Figs 1 and 4) with two nested pairs of incompatible lox sites. In Cre-expressing cells, *ChR2−eYFP* or *eNpHR−eYFP* is first reversibly flipped into the sense orientation via either pair of sites (Fig. 1b); this enables a second irreversible excision that prevents further inversion (Fig. 1b). Double-floxed reversed *ChR2–eYFP* or *eNpHR–eYFP* virus was injected stereotactically into infralimbic cortex of 5–6-week-old *PV::Cre* mice.

***In vivo* recording.** Simultaneous optical stimulation and electrical recording in infralimbic and prelimbic cortex of PV::Cre transgenic mice was carried out as described previously[6]. We recorded local field potentials in mice anesthetized with a ketamine/xylazine mixture, and calculated power spectra immediately after blue light flashes that were separated from other flashes by at least 25 ms.

***In vitro* electrophysiology.** Patch-clamp recordings in brain slices and dynamic clamp were carried out as described previously[5]. All recordings were at $32.5 \pm 1\ ^{\circ}\mathrm{C}$ unless stated otherwise. We stimulated ChR2 in pyramidal neurons using flashes of light lasting 1 ms generated by a 300 W Xenon lamp and a GFP filter set delivered to the slice through a $\times 40$ objective. We used 5-ms light flashes to stimulate ChR2 in fast-spiking interneurons. We stimulated eNpHR in PV interneurons *in vitro* using continuous yellow light generated analogously using a 593 nm filter set.

**sEPSC and light pulse trains.** Experiments were divided into 5- or 10-s long sweeps, which were subdivided into 500-ms segments, containing 0–250 sEPSCs. Rhythmic and non-rhythmic trains used the same sequence of sEPSC rates. For non-rhythmic trains, sEPSCs occurred randomly across each segment. Within rhythmic trains, sEPSCs occurred only during half of each 25 or 125 ms cycle. All summary plots, bar graphs and numerical values in the text are presented as mean $\pm$ s.e.m.

1. Kawaguchi, Y. & Kubota, Y. Neurochemical features and synaptic connections of large physiologically-identified GABAergic cells in the rat frontal cortex. *Neuroscience* **85**, 677–701 (1998).

2. Toledo-Rodriguez, M. *et al.* Correlation maps allow neuronal electrical properties to be predicted from single-cell gene expression profiles in rat neocortex. *Cereb. Cortex* **14**, 1310–1327 (2004).

3. Freund, T. F. Interneuron diversity series: Rhythm and mood in perisomatic inhibition. *Trends Neurosci.* **26**, 489–495 (2003).

4. Whittington, M. A., Traub, R. D. & Jefferys, J. G. Synchronized oscillations in interneuron networks driven by metabotropic glutamate receptor activation. *Nature* **373**, 612–615 (1995).

5. Ylinen, A. *et al.* Intracellular correlates of hippocampal theta rhythm in identified pyramidal cells, granule cells, and basket cells. *Hippocampus* **5**, 78–90 (1995).

6. Tamas, G., Buhl, E. H., Lorincz, A. & Somogyi, P. Proximally targeted GABAergic synapses and gap junctions synchronize cortical interneurons. *Nature Neurosci.* **3**, 366–371 (2000).

7. Fuchs, E. C. *et al.* Recruitment of parvalbumin-positive interneurons determines hippocampal function and associated behavior. *Neuron* **53**, 591–604 (2007).

8. Konig, P., Engel, A. K. & Singer, W. Integrator or coincidence detector? The role of the cortical neuron revisited. *Trends Neurosci.* **19**, 130–137 (1996).

9. Womelsdorf, T. *et al.* Modulation of neuronal interactions through neuronal synchronization. *Science* **316**, 1609–1612 (2007).

10. Lewis, D. A., Hashimoto, T. & Volk, D. W. Cortical inhibitory neurons and schizophrenia. *Nature Rev. Neurosci.* **6**, 312–324 (2005).

11. Orekhova, E. V. *et al.* Excess of high frequency electroencephalogram oscillations in boys with autism. *Biol. Psychiatry* **62**, 1022–1029 (2007).

12. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neurosci.* **8**, 1263–1268 (2005).

13. Zhang, F. *et al.* Multimodal fast optical interrogation of neural circuitry. *Nature* **446**, 633–639 (2007).

14. Aravanis, A. M. *et al.* An optical neural interface: *in vivo* control of rodent motor cortex with integrated fiberoptic and optogenetic technology. *J. Neural Eng.* **4**, S143–S156 (2007).

15. Meyer, A. H., Katona, I., Blatow, M., Rozov, A. & Monyer, H. *In vivo* labeling of parvalbumin-positive interneurons and analysis of electrical coupling in identified neurons. *J. Neurosci.* **22**, 7055–7064 (2002).

16. Livet, J. *et al.* Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007).

17. Conde, F., Lund, J. S. & Lewis, D. A. The hierarchical development of monkey visual cortical regions as revealed by the maturation of parvalbumin-immunoreactive neurons. *Brain Res.* **96**, 261–276 (1996).

18. Raghavachari, S. *et al.* Gating of human theta oscillations by a working memory task. *J. Neurosci.* **21**, 3175–3183 (2001).

19. Howard, M. W. *et al.* Gamma oscillations correlate with working memory load in humans. *Cereb. Cortex* **13**, 1369–1374 (2003).

20. de Ruyter van Stevenick, R. R., Lewen, G. D., Strong, S. P., Koberle, R. & Bialek, W. Reproducibility and variability in neural spike trains. *Science* **275**, 1805–1808 (1997).

21. Arenkiel, B. R. *et al. In vivo* light-induced activation of neural circuitry in transgenic mice expressing channelrhodopsin-2. *Neuron* **54**, 205–218 (2007).

22. Wang, H. *et al.* High-speed mapping of synaptic connectivity using photostimulation in Channelrhodopsin-2 transgenic mice. *Proc. Natl Acad. Sci. USA* **104**, 8143–8148 (2007).

23. Gradinaru, V. *et al.* Targeting and readout strategies for fast optical neural control *in vitro* and *in vivo. J. Neurosci.* **27**, 14231 (2007).

24. Bannister, A. P. Inter- and intra-laminar connections of pyramidal cells in the neocortex. *Neurosci. Res.* **53**, 95–103 (2005).

25. Sanchez-Vives, M. V. & McCormick, D. A. Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nature Neurosci.* **3**, 1027–1034 (2000).

26. Luczak, A., Bartho, P., Marguet, S. L., Buzsaki, G. & Harris, K. D. Sequential structure of neocortical spontaneous activity *in vivo. Proc. Natl Acad. Sci. USA* **104**, 347–352 (2007).

27. Baccus, S. A. & Meister, M. Fast and slow contrast adaptation in retinal circuitry. *Neuron* **36**, 909–919 (2002).

28. Szabadics, J. *et al.* Excitatory effect of GABAergic axo-axonic cells in cortical microcircuits. *Science* **311**, 233–235 (2006).

29. Fries, P., Reynolds, J. H., Rorie, A. E. & Desimone, R. Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* **291**, 1560–1563 (2001).

30. Rodriguez, E. *et al.* Perception's shadow: long-distance synchronization of human brain activity. *Nature* **397**, 430–433 (1999).

*nature*

# LETTERS

# Mechanism of differential control of NMDA receptor activity by NR2 subunits

Marc Gielen[1], Beth Siegler Retchless[2], Laetitia Mony[1], Jon W. Johnson[2] & Pierre Paoletti[1]

*N*-methyl-D-aspartate (NMDA) receptors (NMDARs) are a major class of excitatory neurotransmitter receptors in the central nervous system. They form glutamate-gated ion channels that are highly permeable to calcium and mediate activity-dependent synaptic plasticity[1]. NMDAR dysfunction is implicated in multiple brain disorders, including stroke, chronic pain and schizophrenia[2]. NMDARs exist as multiple subtypes with distinct pharmacological and biophysical properties that are largely determined by the type of NR2 subunit (NR2A to NR2D) incorporated in the heteromeric NR1/NR2 complex[1,3,4]. A fundamental difference between NMDAR subtypes is their channel maximal open probability ($P_o$), which spans a 50-fold range from about 0.5 for NR2A-containing receptors to about 0.01 for receptors containing NR2C and NR2D; NR2B-containing receptors have an intermediate value (about 0.1)[5–9]. These differences in $P_o$ confer unique charge transfer capacities and signalling properties on each receptor subtype[4,6,10,11]. The molecular basis for this profound difference in activity between NMDAR subtypes is unknown. Here we show that the subunit-specific gating of NMDARs is controlled by the region formed by the NR2 amino-terminal domain (NTD), an extracellular clamshell-like domain previously shown to bind allosteric inhibitors[12–15], and the short linker connecting the NTD to the agonist-binding domain (ABD). The subtype specificity of NMDAR $P_o$ largely reflects differences in the spontaneous (ligand-independent) equilibrium between open-cleft and closed-cleft conformations of the NR2-NTD. This NTD-driven gating control also affects pharmacological properties by setting the sensitivity to the endogenous inhibitors zinc and protons. Our results provide a proof of concept for a drug-based bidirectional control of NMDAR activity by using molecules acting either as NR2-NTD 'closers' or 'openers' promoting receptor inhibition or potentiation, respectively.

We first explored the role of the NR2-NTD in the difference of $P_o$ between NR1/NR2A and NR1/NR2B receptors by evaluating the effect of deleting the entire NR2-NTD on receptor activity. We estimated $P_o$ by using a method based on the covalent modification of a cysteine residue introduced in the NR1 subunit (NR1-A652C), which locks open the NMDAR channel[16]. Although this method does not give access to the absolute $P_o$ of receptors containing the wild-type NR1 (NR1wt) subunit, it can report relative differences in channel activity[17]. Indeed, the extent to which the thiol-modifying reagent 2-aminoethylmethanethiosulphonatehydrobromide (MTSEA) potentiates NMDAR currents is inversely related to the channel $P_o$ (ref. 17). MTSEA potentiated currents carried by NR1-A652C/NR2B receptors to a much greater extent than currents of NR1-A652C/NR2A receptors (Fig. 1a, d), consistent with the much lower $P_o$ of NR2B-containing receptors than that of NR2A-containing receptors[5,6,17]. In contrast, MTSEA-induced potentiations of NR1-A652C/NR2A-ΔNTD

and NR1-A652C/NR2B-ΔNTD receptors were indistinguishable (Fig. 1b, d), indicating equal receptor activities. However, receptors incorporating chimaeric NR2A-(2B NTD) or NR2B-(2A NTD) subunits displayed MTSEA-induced potentiations similar to those of the parental NR2 subunits, indicating that swapping the NTDs alone did not exchange the $P_o$ (Fig. 1d). We therefore swapped both the NTD and the highly divergent short (14 residues) linker segment that connects the NTD to the ABD (Supplementary Fig. 1). NR1-A652C/NR2A-(2B NTD+L) and NR1-A652C/NR2B-(2A NTD+L) responses supported levels of MTSEA potentiation closer to those of NR2Bwt-containing and NR2Awt-containing receptors, respectively (Fig. 1c, d). Direct measurement of channel activity with single-channel recordings confirmed this exchange of $P_o$ (Fig. 1e and Supplementary Fig. 2).

We next extended the analysis to the NR2D subunit. MTSEA-induced potentiations of NR2D-containing receptors were considerable (about 300-fold), reflecting the very low $P_o$ of NR1/NR2D receptors (Fig. 1d). Deleting the NR2D-NTD resulted in a fourfold decrease in MTSEA potentiation, indicative of a markedly increased $P_o$ (Fig. 1d). This gain-of-function phenotype could be reinforced by grafting onto NR2D-ΔNTD the NTD plus linker (NTD+L) region of the high-$P_o$ subunit NR2A. Conversely, receptors containing the chimaeric NR2A-(2D NTD+L) subunit displayed 17-fold higher potentiation by MTSEA than NR2Awt-containing receptors, suggestive of a much lower $P_o$ (Fig. 1d). Thus, the low $P_o$ of the NR2D-containing receptors is also set by the NR2-NTD.

Because the estimation of $P_o$ with MTSEA relies on a mutated NR1 subunit (NR1-A652C), we checked that the effects observed did not depend on this mutation. We used the time constant of inhibition by MK-801, an NMDAR open-channel blocker, as an alternative method of assessing $P_o$ (refs 5, 18). Consistent with the higher $P_o$ of receptors containing NR2A than that of receptors containing NR2B, MK-801 inhibited wild-type NR1/NR2A receptors significantly faster than wild-type NR1/NR2B receptors (Supplementary Fig. 3a, b). Deleting the NR2-NTDs abolished this difference (Supplementary Fig. 3b). Whereas swapping the NR2-NTD alone did not exchange MK-801 time constants, incorporating the NTD-ABD linker achieved almost complete transfer (Supplementary Fig. 3a, b). As expected, the onset of MK-801 inhibition at wild-type NR1/NR2D receptors was much slower than that at receptors containing NR2A or NR2B. Deleting the NR2D-NTD or replacing the NTD+L region of NR2D by that of NR2A strongly accelerated MK-801 inhibition, indicative of a much increased $P_o$ (Fig. 1f and Supplementary Fig. 3c). Conversely, MK-801 inhibition of receptors incorporating NR2A-(2D NTD+L) was 15-fold slower than at NR2Awt-containing receptors (Supplementary Fig. 3b). Together with the MTSEA experiments, these results demonstrate that the NR2-NTD+L region is a major determinant of the NR2 subunit-specific activity of NMDARs.

[1]Laboratoire de Neurobiologie, École Normale Supérieure, CNRS, 46 rue d'Ulm, 75005 Paris, France. [2]Department of Neuroscience, University of Pittsburgh, A210 Langely Hall, Pittsburgh, Pennsylvania 15260, USA.
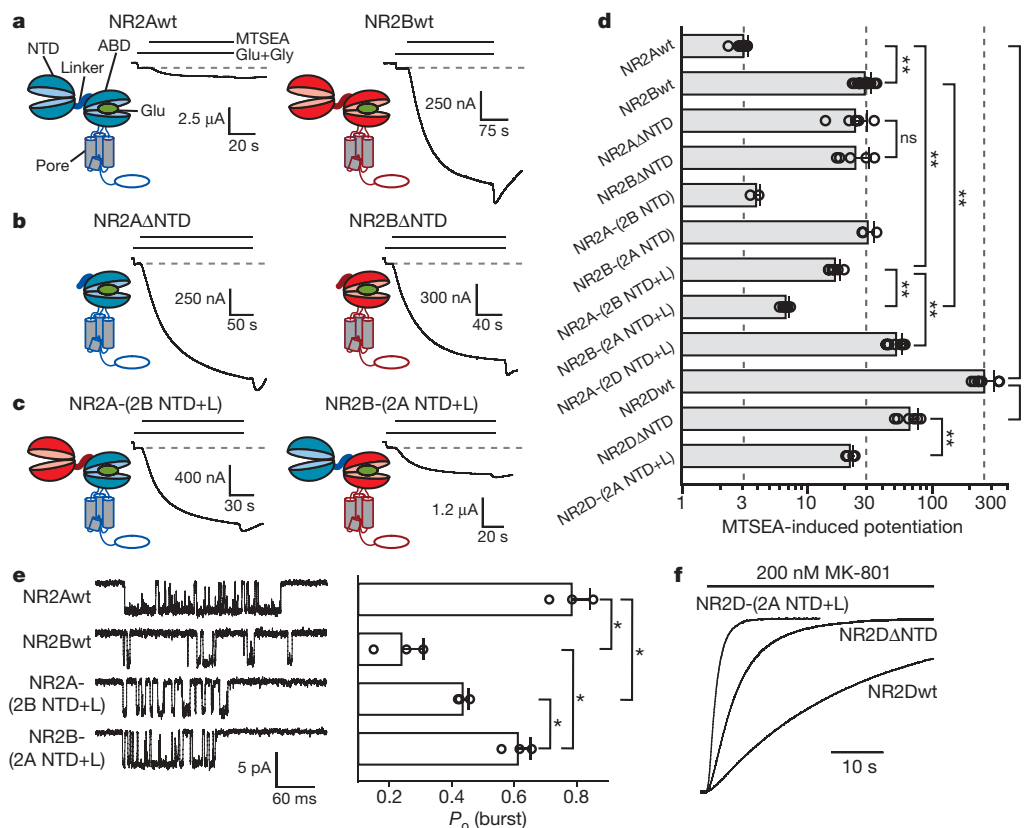
**Figure 1 | The NR2 NTD+L region controls NMDAR $P_o$. a–c,** Potentiation by MTSEA of receptors incorporating NR1-A652C and NR2Awt or NR2Bwt (**a**), NR2A-ΔNTD or NR2B-ΔNTD (**b**), and NR2A-(2B NTD+L) or NR2B-(2A NTD+L) (**c**). **d,** Pooled data (means ± s.d.), from top to bottom: 3.2 ± 0.3 (*n* = 12), 30 ± 4 (*n* = 14), 25 ± 6 (*n* = 6), 25 ± 7 (*n* = 5), 4.0 ± 0.3 (*n* = 3), 32 ± 4 (*n* = 3), 17 ± 2 (*n* = 6), 6.9 ± 0.5 (*n* = 5), 53 ± 7 (*n* = 9), 270 ± 60 (*n* = 7), 68 ± 12 (*n* = 6) and 23 ± 2 (*n* = 5). Two asterisks,

$P < 0.001$. **e,** $P_o$ within bursts of openings for receptors incorporating NR1wt and the indicated NR2 subunit. Left: representative traces of bursts. Right (from top to bottom): 0.78 ± 0.06 (*n* = 3), 0.24 ± 0.07 (*n* = 3), 0.43 ± 0.02 (*n* = 3) and 0.61 ± 0.04 (*n* = 3). Asterisk, $P < 0.05$, Student's *t*-test. Error bars represent s.d. **f,** Kinetics of inhibition by 200 nM MK-801 at receptors incorporating NR1wt and NR2Dwt ($\tau_{on}$ = 32 s), NR2D-ΔNTD (5.7 s) or NR2D-(2A NTD+L) (1.6 s).

We then investigated the mechanism by which a distal domain, the NR2-NTD, influences channel activity. Previous studies on allosteric inhibition of NMDARs by NR2-NTD ligands, such as zinc and

ifenprodil, suggested that these ligands bind the NTD cleft and promote its closure[12,15,19]. This in turn leads to receptor inhibition through disruption of the NR1/NR2 ABD dimer interface, resembling
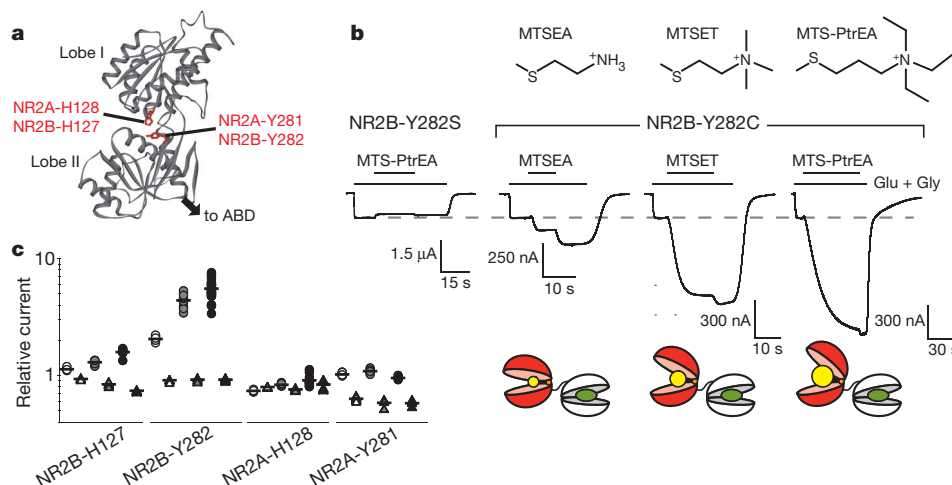


**Figure 2 | Locking open the NR2-NTD increases NMDAR activity. a,** Three-dimensional model of NR2B-NTD. **b,** Top: chemical formula of the transferable moiety of MTSEA, MTSET and MTS-PtrEA. Middle: recordings from NR1wt/NR2B-Y282C and control NR1wt/NR2B-Y282S receptors during treatment with MTS. The potentiation after washout of MTS probably reflects the washout of a reversible pore-blocking effect of the positively

charged MTS. Bottom: schematic representations of the NTD-ABD tandem of NR2B-Y282C after modification by MTS (MTS headgroup in yellow). **c,** Relative currents after application of MTSEA (white symbols), MTSET (grey symbols) and MTS-PtrEA (black symbols) to receptors incorporating NR1wt and the indicated NR2 subunit harbouring a cysteine mutation (circles) or a control mutation (triangles). See Supplementary Table 1 for values.

**704**

the mechanism underlying the desensitization of AMPA ($\alpha$-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptors[20–22]. Because the NTD can adopt at least two conformations, a ligand-free open state and a ligand-bound closed state, we speculated that the NTD-driven control of $P_o$ might result from spontaneous oscillations of the NR2-NTD between an open-cleft conformation, favouring channel opening, and a closed-cleft conformation, favouring pore closure. Similar ligand-independent oscillations have been observed in several clamshell-like proteins, including the bacterial maltose-binding protein[23] (MBP) and the GABA$_B$ ($\gamma$-aminobutyric acid B) receptor[24]. To test this hypothesis, we introduced cysteine residues into the NR2-NTD cleft to lock open the NR2-NTDs with the use of thiol-reactive methanethiosulphonate (MTS) reagents. On the basis of three-dimensional models, we first introduced a cysteine residue deep in the cleft of the NR2B-NTD by mutating the hinge residue NR2B-Y282, whose side chain points towards the cleft entrance[25] (Fig. 2a). Application of the positively charged MTSEA potentiated NR1wt/NR2B-Y282C receptors but not control NR1wt/NR2B-Y282S receptors (Fig. 2b). Using MTS compounds of the same valence but different sizes (2-[trimethylammonium]ethylmethanethiosulphonatebromide [MTSET] and 3-[triethylammonium]propylmethanethiosulphonatebromide [MTS-PtrEA]), we observed that the potentiation increased with increasing MTS size (Fig. 2b, c). Comparison of the rates of inhibition by MK-801 before and after treatment with MTS, together with direct measurement of single-channel activity, revealed that current potentiations reflected an increase in $P_o$ (Supplementary Figs 4 and 5). Sensitivity to glycine (binding the NR1-ABD) was unaltered by treatment with MTS, whereas sensitivity to glutamate (binding the NR2-ABD) was slightly decreased (Supplementary Fig. 6), as expected from the known allosteric interaction between the NR2 NTD and ABD[26]. MTS action was significantly faster on resting receptors than on activated receptors (Supplementary Fig. 7), further arguing for a facilitated opening of the NR2-NTD when the ABD is open. Taken together, these results show that trapping open the NR2-NTD enhances receptor activity. They also indicate that the NTD of NR2B-Y282C is neither permanently open (because there was a potentiating effect of the MTS compounds) nor closed (because the introduced cysteine residue was accessible to MTS), but rather alternates between open and closed conformations, the latter favouring pore closure.

Because NR2B-Y282 is a large residue, we considered the possibility that its mutation into a small residue (cysteine or serine) might have artificially increased the flexibility of the NTD hinge, favouring NTD closure. Indeed, such mutations strongly decreased receptor activity (Supplementary Fig. 8). This effect highlights the unsuspected role of

the NR2-NTD hinge in shaping NMDAR $P_o$, reminiscent of the critical role of the MBP hinge in controlling the apparent maltose affinity[27]. To extend our conclusion of spontaneous NR2-NTD oscillations to receptors with unaltered gating properties, we targeted H127 of NR2B-NTD, because its mutation into cysteine minimally affects receptor activity (Supplementary Fig. 8). MTS compounds still potentiated NR1wt/NR2B-H127C receptors (but not control NR1wt/NR2B-H127A receptors) in a size-dependent manner. However, potentiations were considerably smaller than with NR1wt/NR2B-Y282C receptors (Fig. 2c and Supplementary Fig. 9a). Two reasons may explain this difference: higher basal $P_o$ of NR1wt/NR2B-H127C receptors, and wider opening of the NTD at MTS-modified NR2B-Y282C subunits because of the deeper location of Y282 in the cleft. Taken together, these results provide the new information that spontaneous oscillations of the NR2B-NTD contribute to the low $P_o$ of wild-type NR1/NR2B receptors.

We then tested the prediction that the high $P_o$ of NR2A-containing receptors results from the preference of NR2A-NTD for the open conformation. As with NR2B, we found the $P_o$ of NR2A-containing receptors to be significantly decreased by the mutation of NR2A-Y281 into small residues (Supplementary Fig. 8). A potentiating component was also observed at receptors containing NR2A-Y281C during treatment with MTS compounds, but not at control NR2A-Y281A receptors. However, MTS-induced potentiations were much smaller than at NR2B-Y282C receptors and were independent of MTS size (Fig. 2c and Supplementary Fig. 9b), suggesting that the NR2A-NTD is much less sensitive to steric hindrance than the NR2B-NTD. In addition, no potentiation was observed at NR2A-H128C receptors even with the larger MTSET and MTS-PtrEA (Fig. 2c and Supplementary Fig. 9b). This is consistent with the idea that NR2A-NTD spends most of its time in an open-cleft conformation, thus contributing to the relatively high $P_o$ of NR2A-containing receptors.

Our results on chimaeric NR2 subunits, showing that the NTD-ABD linker is required for the differential influence of the NR2-NTD on receptor $P_o$, raised the possibility that this element is also crucial during the allosteric modulation of NMDARs by NTD ligands. NR2A-NTD forms a high-affinity zinc inhibitory site[12–14]; in accord with this, NR1wt/NR2D-(2A NTD+L) receptors were highly sensitive to zinc (Fig. 3a). NR1wt/NR2B-(2A NTD) receptors are also highly sensitive to zinc. Surprisingly, zinc is much more potent at these receptors than at wild-type NR1/NR2A receptors (Fig. 3a), suggesting that the NR2B NTD-ABD linker facilitates NTD-cleft closure. Increasing the chimaera length to incorporate the NR2A NTD-ABD linker almost completely restored NR2Awt-like zinc sensitivity (Fig. 3a). This again highlights the importance of the
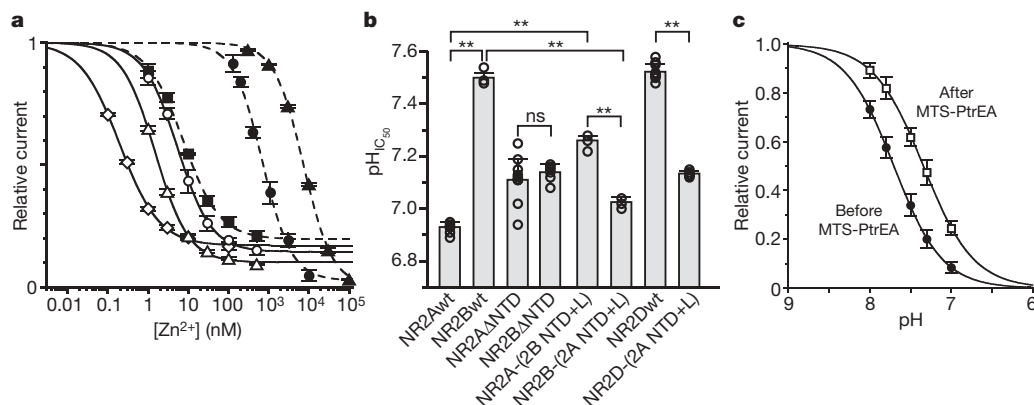


**Figure 3 | The NR2 NTD+L region controls zinc and proton sensitivities of NMDARs. a**, Zinc sensitivity of receptors incorporating NR1wt and NR2Awt (filled squares; maximum inhibition 81%, IC$_{50}$ 7.5 nM; $n = 6$), NR2Bwt (filled circles; 98%, 720 nM; $n = 13$), NR2Dwt (filled triangles; 100%, 7.8 $\mu$M; $n = 3$), NR2B-(2A NTD) (open diamonds; 83%, 0.20 nM; $n = 4$), NR2B-(2A NTD+L) (open circles; 86%, 5.4 nM; $n = 4$) or NR2D-

(2A NTD+L) (open triangles; 90%, 1.5 nM; $n = 5$). $n_H$ was in the range 0.9–1.2. **b**, pH$_{IC_{50}}$ of receptors incorporating NR1wt and the indicated NR2 subunits. See Supplementary Table 2 for values. Two asterisks, $P < 0.001$. **c**, Proton sensitivity of NR1wt/NR2B-Y282C receptors before (pH$_{IC_{50}}$ = 7.70, $n_H$ = 1.5; $n = 3$) and after (pH$_{IC_{50}}$ = 7.34, $n_H$ = 1.4; $n = 3$) modification by MTS-PtrEA. Error bars represent s.d.

NTD-ABD linker for communication between the NTD and the gating machinery.

Protons are another allosteric modulator that differentially inhibit NMDAR subtypes[1]. In contrast with the zinc sensor, the proton sensor is thought to be closely associated with the channel gate[28]. Unexpectedly, deleting the NR2-NTDs fully abolished the difference in pH sensitivity between wild-type NR1/NR2A and NR1/NR2B receptors (Fig. 3b). Moreover, swapping the NTD+L region between NR2A and NR2B reversed their pH sensitivities, whereas grafting the NR2A NTD+L region onto NR2D decreased its proton sensitivity towards that of NR2Awt-containing receptors (Fig. 3b). Proton sensitivity was also decreased when the NR2B-Y282C NTD was locked open with MTS-PtrEA (Fig. 3c). Therefore the NR2 dependence of pH sensitivity is unlikely to result from an intrinsic difference in the proton sensor between NR1/NR2 receptor subtypes, but rather from differential access to the proton-binding site owing to the influence of NR2-NTD on channel activity.

Our study reveals that the large differences in channel activity conferred by the various NR2 NMDAR subunits originate from a region remote from the agonist-binding/channel gating core. This region comprises the large NR2-NTD and the short linker connecting the NR2-NTD to the ABD. The bilobate NR2-NTD oscillates spontaneously between open-cleft and closed-cleft conformations (Fig. 4), the latter triggering disruption of the ABD dimer interface and subsequent channel closure[20]. The NTD-ABD linker could exert its key influence by tuning the equilibrium between the different conformations of the NR2-NTD. The identity of the NR2-NTD+L region also determines the sensitivity to zinc and protons, two endogenous allosteric inhibitors of NMDARs that are likely to be critical in the regulation of NMDAR activity under physiological and pathological conditions[1,3]. Through its dynamic conformational equilibrium, the NR2-NTD could serve as a target for either negative or positive subunit-specific allosteric modulators (Fig. 4). Compounds such as ifenprodil, which bind the NTD cleft and promote its closure (NTD 'closers'), behave as subunit-specific NMDAR inhibitors and show good efficacy as neuroprotectants[2]. We propose that molecules that bind the same cleft but impede its closure (NTD 'openers') would behave as NMDAR potentiators (Fig. 4). Such molecules may prove to be of significant therapeutic benefit, given the accumulating evidence that major human psychoses, including schizophrenia, are associated with a deficit of NMDAR activity[2,29].

## METHODS SUMMARY

**Complementary DNA constructs and site-directed mutagenesis.** The pcDNA3-based expression plasmids, mutagenesis and sequencing procedure have been described previously[19]. Chimaeras were obtained by classical amplification by polymerase chain reaction and subsequent subcloning into the parental clone.

**Electrophysiology.** Recombinant NMDARs were expressed in *Xenopus laevis* oocytes after simultaneous injection of cDNAs (at 10 ng μl$^{-1}$; nuclear injection) coding for the various NR1 and NR2 subunits (ratio 1:1). Oocytes were prepared, injected, voltage-clamped and superfused as described previously[12].

Single channels were recorded from human embryonic kidney (HEK)-cell outside-out patches.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Dingledine, R., Borges, K., Bowie, D. & Traynelis, S. F. The glutamate receptor ion channels. *Pharmacol. Rev.* **51**, 7–61 (1999).
2. Kemp, J. A. & McKernan, R. M. NMDA receptor pathways as drug targets. *Nature Neurosci.* **5** (Suppl.), 1039–1042 (2002).
3. Paoletti, P. & Neyton, J. NMDA receptor subunits: function and pharmacology. *Curr. Opin. Pharmacol.* **7**, 39–47 (2007).
4. Cull-Candy, S. G. & Leszkiewicz, D. N. Role of distinct NMDA receptor subtypes at central synapses. *Sci. STKE* **2004**, re16 (2004).
5. Chen, N., Luo, T. & Raymond, L. A. Subtype-dependence of NMDA receptor channel open probability. *J. Neurosci.* **19**, 6844–6854 (1999).
6. Erreger, K., Dravid, S. M., Banke, T. G., Wyllie, D. J. & Traynelis, S. F. Subunit-specific gating controls rat NR1/NR2A and NR1/NR2B NMDA channel kinetics and synaptic signalling profiles. *J. Physiol. (Lond.)* **563**, 345–358 (2005).
7. Wyllie, D. J., Behe, P. & Colquhoun, D. Single-channel activations and concentration jumps: comparison of recombinant NR1a/NR2A and NR1a/NR2D NMDA receptors. *J. Physiol. (Lond.)* **510**, 1–18 (1998).
8. Dravid, S. M., Prakash, A. & Traynelis, S. F. Activation of recombinant NR1/NR2C NMDA receptors. *J. Physiol. (Lond.)* **586**, 4425–4439 (2008).
9. Popescu, G., Robert, A., Howe, J. R. & Auerbach, A. Reaction mechanism determines NMDA receptor response to repetitive stimulation. *Nature* **430**, 790–793 (2004).
10. Liu, Y. *et al.* NMDA receptor subunits have differential roles in mediating excitotoxic neuronal death both *in vitro* and *in vivo. J. Neurosci.* **27**, 2846–2857 (2007).
11. Liu, L. *et al.* Role of NMDA receptor subtypes in governing the direction of hippocampal synaptic plasticity. *Science* **304**, 1021–1024 (2004).
12. Paoletti, P. *et al.* Molecular organization of a zinc binding N-terminal modulatory domain in a NMDA receptor subunit. *Neuron* **28**, 911–925 (2000).
13. Low, C. M., Zheng, F., Lyuboslavsky, P. & Traynelis, S. F. Molecular determinants of coordinated proton and zinc inhibition of N-methyl-D-aspartate NR1/NR2A receptors. *Proc. Natl Acad. Sci. USA* **97**, 11062–11067 (2000).
14. Choi, Y. B. & Lipton, S. A. Identification and mechanism of action of two histidine residues underlying high-affinity Zn$^{2+}$ inhibition of the NMDA receptor. *Neuron* **23**, 171–180 (1999).
15. Perin-Dureau, F., Rachline, J., Neyton, J. & Paoletti, P. Mapping the binding site of the neuroprotectant ifenprodil on NMDA receptors. *J. Neurosci.* **22**, 5955–5965 (2002).
16. Jones, K. S., VanDongen, H. M. & VanDongen, A. M. The NMDA receptor M3 segment is a conserved transduction element coupling ligand binding to channel opening. *J. Neurosci.* **22**, 2044–2053 (2002).
17. Yuan, H., Erreger, K., Dravid, S. M. & Traynelis, S. F. Conserved structural and functional control of N-methyl-D-aspartate receptor gating by transmembrane domain M3. *J. Biol. Chem.* **280**, 29708–29716 (2005).
18. Blanke, M. L. & VanDongen, A. M. Constitutive activation of the N-methyl-D-aspartate receptor via cleft-spanning disulfide bonds. *J. Biol. Chem.* **283**, 21519–21529 (2008).
19. Rachline, J., Perin-Dureau, F., Le Goff, A., Neyton, J. & Paoletti, P. The micromolar zinc-binding domain on the NMDA receptor subunit NR2B. *J. Neurosci.* **25**, 308–317 (2005).
20. Gielen, M. *et al.* Structural rearrangements of NR1/NR2A NMDA receptors during allosteric inhibition. *Neuron* **57**, 80–93 (2008).
21. Sun, Y. *et al.* Mechanism of glutamate receptor desensitization. *Nature* **417**, 245–253 (2002).
22. Mayer, M. L. Glutamate receptors at atomic resolution. *Nature* **440**, 456–462 (2006).
23. Tang, C., Schwieters, C. D. & Clore, G. M. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449**, 1078–1082 (2007).
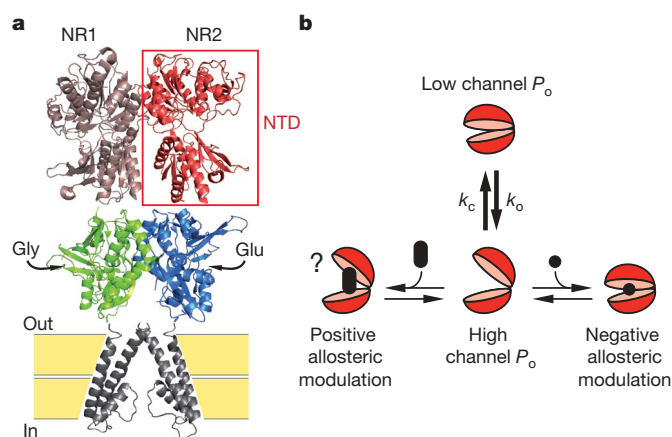


**Figure 4 | Model for the control of NMDAR activity by the N-terminal domain of NR2. a,** Structural depiction of an NMDAR. The full receptor is a tetramer, but only an NR1/NR2 dimer is shown[30]. **b,** In its ligand-free state, the NR2-NTD alternates between open-cleft and closed-cleft conformations, the latter favouring pore closure. In the model, this equilibrium determines the subtype specificity of NMDAR $P_o$; $k_o/k_c$(NR2B) $<$ $k_o/k_c$(NR2A). The NTD is also the target of subunit-specific allosteric inhibitors such as zinc[12–14,19] or ifenprodil[15,25], which bind the central cleft of the NTD and promote domain closure. We propose that a molecule binding in the same cleft, but preventing its closure, behaves as a positive allosteric modulator, enhancing receptor activity.

24. Kniazeff, J. *et al.* Locking the dimeric GABA_B G-protein-coupled receptor in its active state. *J. Neurosci.* **24**, 370–377 (2004).

25. Mony, L. *et al.* Structural basis of NR2B-selective antagonist recognition by *N*-methyl-ᴅ-aspartate receptors. *Mol. Pharmacol.* **75**, 60–74 (2009).

26. Zheng, F. *et al.* Allosteric interaction between the amino terminal domain and the ligand binding domain of NR2A. *Nature Neurosci.* **4**, 894–901 (2001).

27. Marvin, J. S. & Hellinga, H. W. Manipulation of ligand binding affinity by exploitation of conformational coupling. *Nature Struct. Biol.* **8**, 795–798 (2001).

28. Low, C. M. *et al.* Molecular determinants of proton-sensitive *N*-methyl-ᴅ-aspartate receptor gating. *Mol. Pharmacol.* **63**, 1212–1222 (2003).

29. Lisman, J. E. *et al.* Circuit-based framework for understanding neurotransmitter and risk gene interactions in schizophrenia. *Trends Neurosci.* **31**, 234–242 (2008).

30. Furukawa, H., Singh, S. K., Mancusso, R. & Gouaux, E. Subunit arrangement and function in NMDA receptors. *Nature* **438**, 185–192 (2005).

## METHODS

**Two electrode voltage-clamp recordings and analysis.** For all experiments, except those aimed at measuring pH sensitivity, the standard external solution contained (in mM): 100 NaCl, 2.5 KCl, 0.3 BaCl$_2$, 5 HEPES; pH adjusted to 7.3 with NaOH. To chelate trace amounts of contaminant zinc, diethylenetriamine-pentaacetic acid (DTPA) (10 μM) was added to all the '0' zinc solutions[31]. For free zinc concentrations in the range 1 nM–1 μM, tricine (10 mM) was used to buffer zinc, whereas *N*-[2-acetamido]-iminodiacetic acid (ADA) (1 mM) was used to buffer zinc in the 0.1–100 nM range[20]. For the pH sensitivity experiments, an enriched HEPES external solution was used to ensure proper pH buffering[20]. Currents were elicited by the simultaneous application of saturating concentrations of glutamate and glycine (100 μM each), and measured at a holding potential of −60 mV. MTS compounds were used at 0.2 mM (except in Supplementary Fig. 7). Experiments were performed at room temperature (18–25 °C). Data collection and analysis of pH and zinc dose–response curves were performed as described in ref. 20. MK-801 time constants of inhibition were obtained by fitting currents with a single-exponential component within a time window corresponding to 10–90% of the maximal inhibition. Data points used for statistical tests were assumed log-normally distributed before performing a Student's *t*-test (unless otherwise indicated).

**Single-channel recordings and analysis.** HEK cells were transfected with 2 μg of cDNAs mixed in a ratio of 1 NR1:3 NR2:3 green fluorescent protein (GFP), using calcium phosphate precipitation or FuGENE Transfection Reagent (Roche). Positive cells were revealed by GFP epifluorescence. Patch pipettes of 5–10 MΩ were filled with a solution containing (in mM): 115 CsF, 10 CsCl, 10 HEPES, 10 EGTA; pH adjusted to 7.15 with CsOH. The osmolality was 270 mosmol kg$^{-1}$. The standard external solution contained (in mM): 140 NaCl, 2.8 KCl, 0.5 CaCl$_2$, 10 HEPES, 0.01 EDTA; pH adjusted to 7.3 with NaOH. Osmolality was adjusted to 290 mosmol kg$^{-1}$ with sucrose. EDTA was added to chelate trace amounts of contaminant zinc[31]. Channel openings were activated by 100 μM glycine, with 0.05 or 0.01 μM glutamate in most experiments, or with 100 μM glutamate in some patches (included only if no double openings were observed). The holding potential (after correction for junction potential) was −80 to −90 mV. Experiments were performed at room temperature. Currents were recorded with an Axopatch 200B amplifier (Molecular Devices), sampled at 20–50 kHz, low-pass filtered (eight-pole Bessel) at 5–10 kHz. Before analysis of $P_o$ within a burst, data were digitally refiltered to give a cascaded low-pass filter cutoff frequency of 2 kHz. pClamp 9 or 10 (Molecular Devices) was used to acquire and analyse the data.

The principal goal of single-channel analysis was to measure the open probability ($P_o$) within bursts of channel openings, which provides a good estimate of $P_o$ within an NMDAR activation[6,32,33]. To idealize single-channel data, transitions were detected by using a 50% threshold criterion[34]. Events of 200 μs duration or less were excluded from analysis. Missing and ignoring brief events can significantly influence dwell-time histograms. However, such brief events contribute only a tiny fraction of the total time that a channel spends open or closed. Thus, missed events should not have significantly affected measurements of $P_o$. Histograms are presented as square root versus log(time) plots[35]. Shut-time histograms were fitted with three or four exponential components. A burst was defined as a series of openings separated by closures of duration less than a critical duration, $T_{crit}$. Bursts with two levels of openings were discarded. We calculated $T_{crit}$ between the two longest components of the shut-time histograms so that total number of event misclassifications was minimized[34,36]. For NR1wt/NR2Awt and NR1wt/NR2B-(2A NTD+L) receptors, the two longest components of the shut-time distribution differed by a mean factor of more than 390, whereas these components were less separated for NR1wt/NR2Bwt and NR1wt/NR2A-(2B NTD+L) (23-fold and 54-fold separation, respectively). For the latter two constructs, the separation between shut-time components results in a greater than desired number of misclassification of shut times[34]. This may have led to an overestimation of the $P_o$ within a burst. However, for wild-type receptors, our data are consistent overall with previous results[6,33], suggesting that our estimates of $P_o$ are reliable.

**Chemicals.** HEPES, L-glutamate, glycine, DTPA, EDTA, tricine and ADA were obtained from Sigma, D-APV from Ascent Scientific, 2-aminoethylmethanethio-sulphonatehydrobromide (MTSEA), 2-(trimethylammonium)ethylmethanethio-sulphonatebromide (MTSET) and 3-(triethylammonium)propylmethanethio-sulphonatebromide (MTS-PtrEA) from Toronto Research Chemicals, and (+)-MK-801 from Tocris. MTS compounds were prepared as 40 mM stock solutions in doubly distilled water, aliquoted in small volumes (50 μl) and stored at −20 °C; aliquots were thawed just before use.

**Construction of Fig. 4.** The molecular architecture shown in Fig. 4a was illustrated by the crystal structure of the mGluR1 ligand-binding domain dimer (PDB 1ewv, ref. 37) at the level of the NTD, the NMDAR NR1/NR2A agonist-binding domain dimer (PDB 2a5T, ref. 30) and two subunits of the KcsA tetramer (PDB 1bl8, ref. 38) as the transmembrane region of the receptor. The third transmembrane segment and the C-terminal cytoplasmic region are lacking in this structural depiction.

31. Paoletti, P., Ascher, P. & Neyton, J. High-affinity zinc inhibition of NMDA NR1-NR2A receptors. *J. Neurosci.* **17**, 5711–5725 (1997).
32. Erreger, K. & Traynelis, S. F. Zinc inhibition of rat NR1/NR2A *N*-methyl-D-aspartate receptors. *J. Physiol. (Lond.)* **586**, 763–778 (2008).
33. Schorge, S., Elenes, S. & Colquhoun, D. Maximum likelihood fitting of single channel NMDA activity with a mechanism composed of independent dimers of subunits. *J. Physiol. (Lond.)* **569**, 395–418 (2005).
34. Colquhoun, D. & Sigworth, F. J. in *Single-channel Recording* (eds Sakmann, B. & Neher, E.) 483–587 (Plenum, 1995).
35. Sigworth, F. J. & Sine, S. M. Data transformations for improved display and fitting of single-channel dwell time histograms. *Biophys. J.* **52**, 1047–1054 (1987).
36. Jackson, M. B., Wong, B. S., Morris, C. E., Lecar, H. & Christian, C. N. Successive openings of the same acetylcholine receptor channel are correlated in open time. *Biophys. J.* **42**, 109–114 (1983).
37. Kunishima, N. *et al.* Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature* **407**, 971–977 (2000).
38. Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K$^+$ conduction and selectivity. *Science* **280**, 69–77 (1998).

# LETTERS

# Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors

Jun K. Takeuchi[1,2] & Benoit G. Bruneau[1,3]

Heart disease is the leading cause of mortality and morbidity in the western world. The heart has little regenerative capacity after damage, leading to much interest in understanding the factors required to produce new cardiac myocytes. Despite a robust understanding of the molecular networks regulating cardiac differentiation[1,2], no single transcription factor or combination of factors has been shown to activate the cardiac gene program *de novo* in mammalian cells or tissues. Here we define the minimal requirements for transdifferentiation of mouse mesoderm to cardiac myocytes. We show that two cardiac transcription factors, Gata4 and Tbx5, and a cardiac-specific subunit of BAF chromatin-remodelling complexes, Baf60c (also called Smarcd3), can direct ectopic differentiation of mouse mesoderm into beating cardiomyocytes, including the normally non-cardiogenic posterior mesoderm and the extraembryonic mesoderm of the amnion. Gata4 with Baf60c initiated ectopic cardiac gene expression. Addition of Tbx5 allowed differentiation into contracting cardiomyocytes and repression of non-cardiac mesodermal genes. Baf60c was essential for the ectopic cardiogenic activity of Gata4 and Tbx5, partly by permitting binding of Gata4 to cardiac genes, indicating a novel instructive role for BAF complexes in tissue-specific regulation. The combined function of these factors establishes a robust mechanism for controlling cellular differentiation, and may allow reprogramming of new cardiomyocytes for regenerative purposes.

The transcriptional regulation of the developing heart has been well studied[1,2], but the factors sufficient to induce the cardiac program in mammalian cells have remained elusive. Recent work has demonstrated important roles for members of the polymorphic Swi/Snf-like BAF chromatin remodelling complexes in cell-type specification and differentiation[3–7]. Baf60c, a cardiac-enriched BAF complex subunit, physically links DNA-binding transcription factors to BAF complexes, thereby modulating the transcription of target genes[3]. Mouse embryos with reduced levels of Baf60c have severe heart defects and defective cardiac differentiation[3]. Because Baf60c is expressed specifically in precardiac mesoderm, we determined whether it is necessary for the activity of essential cardiac DNA-binding factors in non-cardiac cells.

We transiently transfected cultured mouse embryos with expression constructs for Baf60c and combinations of three transcription factors that are important for activation of cardiac genes[1,2]: the zinc-finger transcription factor Gata4, the homeodomain transcription factor Nkx2-5, and the T-box transcription factor Tbx5 (Fig. 1a). Induction of cardiac differentiation was assessed by expression of the early cardiac marker *Actc1*, encoding alpha cardiac actin. Control transfections with enhanced green fluorescent protein (EGFP)-expressing constructs or Tbx5, Nkx2-5 and Gata4 combined did not induce *Actc1* (Fig. 1a, 0 out of 20 and 0 out of 15 embryos, respectively). In contrast, co-transfection of Tbx5, Nkx2-5 and Gata4

with Baf60c led to markedly expanded and ectopic activation of *Actc1* (Fig. 1a, d, 9 out of 11 embryos). We could reliably induce *Actc1* between embryonic day (E)6.5 and E8.75, but later transfections proved ineffective, indicating a limited time window for *Actc1* induction or technical limitations inherent to our approach. *Actc1* induction was not potentiated by myocardin, a transcriptional coactivator that activates some cardiac genes in cell culture[8], or the precardiac mesoderm transcription factor Mesp1 (ref. 9), which can promote cardiac lineages in embryonic stem cells[10–12].

In embryos co-transfected with Tbx5, Nkx2-5 and Gata4 plus Baf60c, we also detected ectopic *Myl2* messenger RNA, and α-tropomyosin (α-TM; also called Tpm1) and cardiac troponin T (cTnT; also called Tnnt2) proteins, specific markers of the embryonic heart (Figs 1b and 2a). Induction of cardiac markers was confined to transfected cells, suggesting a cell-autonomous effect (Figs 1c and 2a, c). Notably, ectopic beating cardiac myocytes were observed in normally non-cardiogenic mesoderm transfected with Tbx5, Nkx2-5 and Gata4 plus Baf60c (9 out of 16 embryos), suggesting induction of a full cardiac program (Fig. 1d and Supplementary Movie 1). Ectopic contractile tissue was observed even though the endogenous cardiac field was not yet beating, indicating accelerated cardiac differentiation. Thus, a simple combination of DNA-binding transcription factors and the chromatin-remodelling protein Baf60c induced cardiac differentiation *in vivo*.

cTnT was induced predominantly in mesoderm, but not in endoderm (Fig. 2a, b). We could not reliably transfect ectoderm. Ectopic cardiac markers were activated in all mesodermal regions, including the posterior and medial mesoderm, which are normally not fated to cardiac lineages (Figs 1a, 2a, 3a and 4b). Notably, ectopic cTnT and beating tissue could be induced in the amnion, an extraembryonic tissue (Fig. 2c). The amnion is composed of a mesoderm layer derived from the primitive streak[9,13], and an ectodermal layer, suggesting that endodermal signals were dispensable for cardiogenic induction. Thus, non-cardiac mesoderm was directly programmed into cardiac tissue by the combination of Tbx5, Nkx2-5 and Gata4 plus Baf60c.

We sought to define the minimal set of factors required to induce cardiac differentiation. Ectopic *Actc1* was efficiently induced by Gata4 plus Baf60c (9 out of 11 embryos) but not by Nkx2-5 plus Baf60c or Tbx5 plus Baf60c (Figs 3a, c and 4b). *Nkx2-5* was induced by Gata4 plus Baf60c (Figs 3b and 4b), but *Tbx5* was not (not shown). As Gata4 and Nkx2-5 cooperatively activate numerous cardiac genes[1,2], the induction of *Nkx2-5* by Gata4 plus Baf60c provides an important feed-forward mechanism to establish and reinforce the cardiac program.

We substituted the haematopoietic GATA factor Gata1 for Gata4 (ref. 14) and Baf60a or Baf60b, which are not expressed in the heart[3], for Baf60c. Gata1 functionally replaced Gata4 but with reduced potency, and Baf60b (but not Baf60a) replaced Baf60c, again with
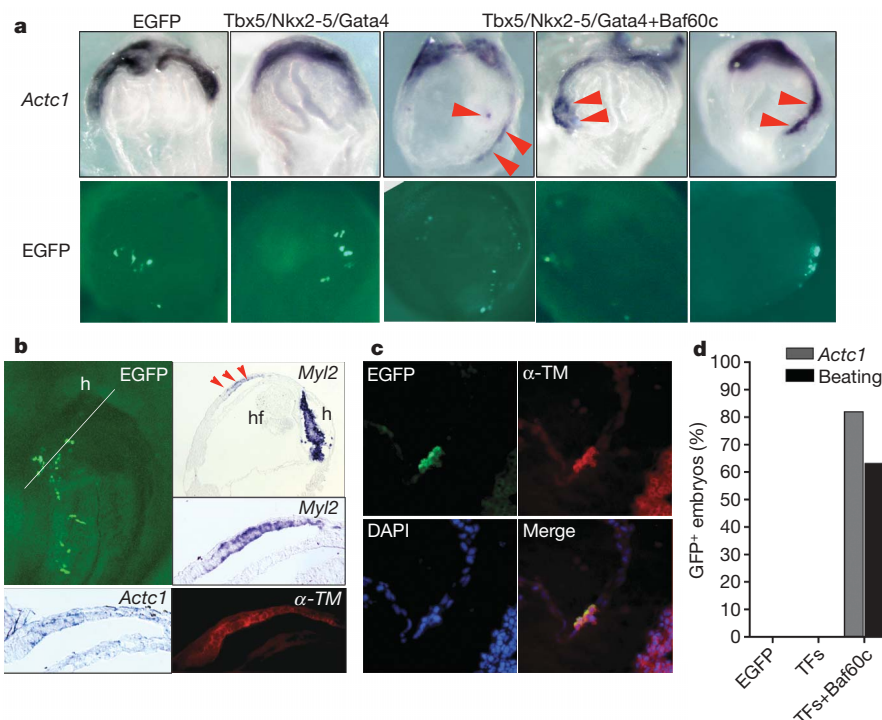
**Figure 1 | Ectopic induction of cardiac differentiation in mouse embryos.**
**a**, *Actc1 in situ* hybridization (top) shows an endogenous cardiac crescent at E8.0 and ectopic cardiac gene expression (red arrowheads) in embryos transfected with the indicated expression constructs. The bottom row shows EGFP signal. **b**, Ectopic expression of *Myl2*, *α-TM* and *Actc1* in consecutive sections of an embryo transfected with Tbx5, Nkx2-5 and Gata4 plus Baf60c. The top-left panel shows EGFP signal and plane of section. h, heart; hf,

headfold. **c**, Ectopic α-TM expression (red) is restricted to EGFP+ cells (green). 4,6-diamidino-2-phenylindole (DAPI) shows nuclei (blue). **d**, Percentage of embryos with ectopic *Actc1*-positive foci and beating tissue. EGFP, 0 out of 20 embryos; TFs, 0 out of 15 embryos; TFs + Baf60c, 9 out of 11 *Actc1*+ embryos and 9 out of 16 beating embryos. TFs, transcription factors (Tbx5, Nkx2-5 and Gata4). Original magnification: **a**, ×10; **b**, ×10 (top left) and ×20; **c**, ×40.
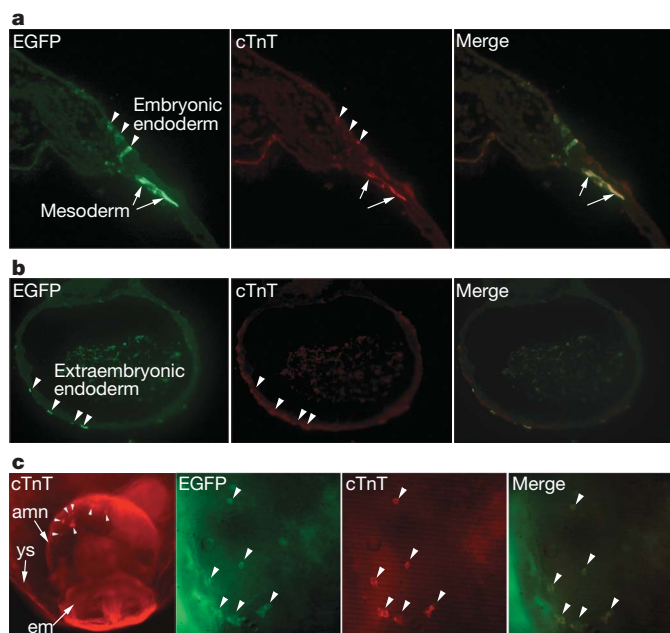


**Figure 2 | Induction of cardiac differentiation in embryonic and extraembryonic mesoderm. a**, Transfected cells were observed in endoderm and mesoderm (EGFP), but induction of cTnT was mainly restricted to mesoderm. **b**, cTnT was not induced in extraembryonic endoderm. **c**, Induction of cTnT in mouse amnion by Gata4 and Tbx5 plus Baf60c. Left panel: low-magnification view of a transfected embryo; arrowheads show ectopic cTnT in the amnion. Right panels: close-up view of the amnion, showing transfected cells (EGFP), cTnT expression and a merge of both. Arrowheads indicate transfected cTnT+ cells. amn, amnion; em, embryo; ys, yolk sac. Original magnification: **a**, ×20; **b**, ×10; **c**, ×5 (left panel) and ×25.

reduced potency (Fig. 3d). Gata1 and Baf60b together were ineffective at activating *Actc1*. These results indicate a significant degree of tissue specificity conferred by Gata4 and Baf60c.

Although Gata4 plus Baf60c initiated cardiac gene expression, they did not induce contractile tissue (Fig. 3c, 0 out of 10). We hypothesized that additional input from Tbx5 was required to complete the cardiac program. To investigate this possibility, we used fluorescence-activated cell sorting (FACS) to isolate EGFP-positive cells from the non-cardiogenic posterior mesoderm of transfected embryos and isolated RNA for polymerase chain reaction with reverse transcription (RT–PCR) analysis. Some cardiac genes were induced by Gata4 plus Baf60c, but Tbx5 was required to induce additional genes required for cardiac function (Fig. 4b, c and Supplementary Fig. 1). Accordingly, transfection of Gata4 and Tbx5 with Baf60c induced ectopic beating tissue (8 out of 16 embryos; Fig. 3c and Supplementary Movie 2). Thus, we defined a minimal set of transcriptional regulators required to induce cardiac differentiation.

*Isl1* expression, which marks early cardiac progenitors[15], was not observed in transfected cells, indicating that we did not target un-differentiated cardiac precursors (Fig. 4b). However, *Isl1* was induced as part of the cardiac program induced by the combination of Gata4, Tbx5 and Baf60c. Interestingly, expression of the posterior mesoderm markers *Tbx6* and *Mesp1* was reduced in cells transfected with the combination of Gata4, Tbx5 and Baf60c, a finding that suggests reprogramming of mesoderm (Fig. 4b).

A combination of tissue-restricted BAF complex subunits may be required to enact a specific transcriptional program. The BAF complex subunits Baf45a and Baf45b have been implicated in neural progenitor- and neuron-specific transcription[5,7]. Baf45c (also known as Dpf3) is predominantly expressed in the developing heart[6] and recognizes specific histone modifications, suggesting that it may enhance the specificity of target gene recognition[6]. Expression constructs for Baf45c alone or with Tbx5, Nkx2-5 and Gata4 did not
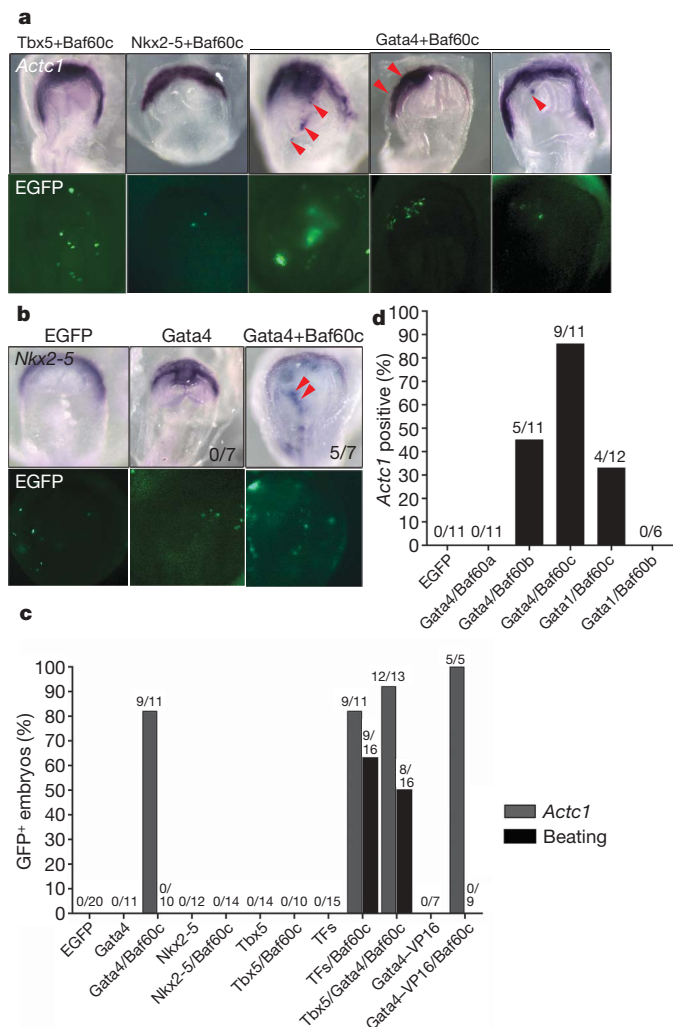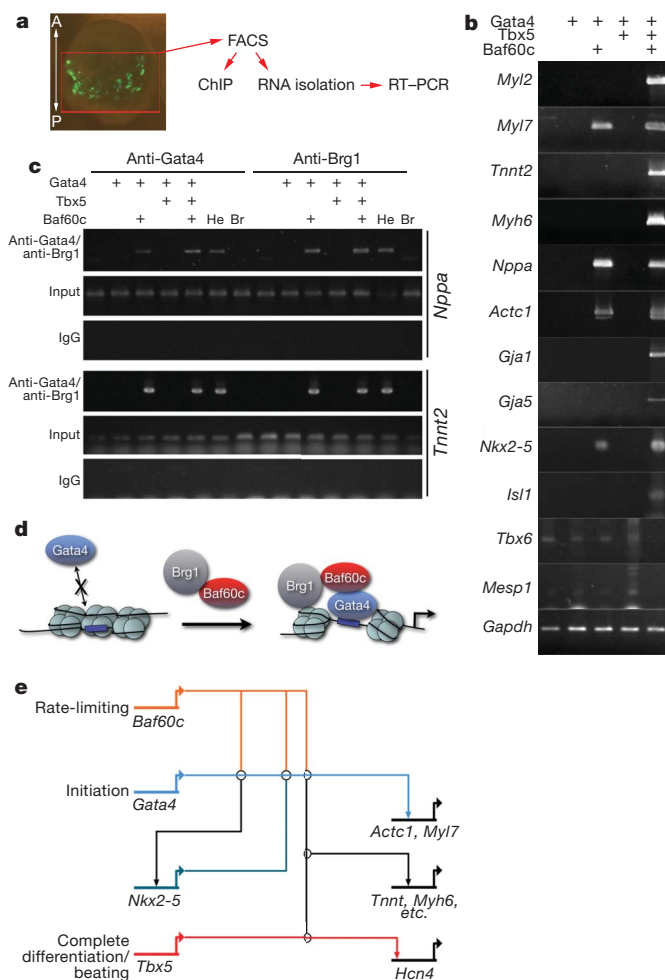
**Figure 3 | Minimal requirements for cardiac gene activation in mouse embryos. a**, Expression of *Actc1* in embryos transfected with pairwise combinations of transcription factors plus Baf60c. Only Gata4 plus Baf60c induced ectopic *Actc1* (red arrowheads). **b**, Expression of *Nkx2-5*; Gata4 plus Baf60c induced ectopic *Nkx2-5* (arrowheads). **c**, Percentage of *Actc1*-positive and beating embryos for each combination. TFs, transcription factors (Tbx5, Nkx2-5 and Gata4). **d**, Specificity of cardiogenic factors; percentage of embryos in which *Actc1* was induced is shown. In **c** and **d**, numbers indicate number of positive embryos per number of experimental embryos. Original magnification: **a**, **b**, ×10.



**Figure 4 | Mechanism for induction of cardiac differentiation. a**, Strategy for isolating and analysing transfected cells. A, anterior; P, posterior. **b**, RT–PCR of several cardiac markers in RNA isolated from transfected EGFP-positive mouse embryonic cells obtained by FACS. **c**, Chromatin immunoprecipitation shows that Gata4 and Brg1 bind *Tnnt2* and *Nppa* only in the presence of Baf60c. Br, brain; He, heart; IgG, nonspecific immunoserum. **d**, Model for action of Baf60c. **e**, Minimal transcriptional network for the ectopic induction of cardiac differentiation. (This panel was made using Biotapestry software.)

induce cardiac gene expression (0 out of 8), and the addition of Baf45c to Tbx5, Nkx2-5 and Gata4 plus Baf60c did not increase the efficiency of induction (5 out of 6 versus 5 out of 6).

We investigated the mechanism by which Baf60c potentiates the function of Gata4. If Baf60c functions primarily to promote transcriptional activity of Gata4 independent of chromatin remodelling, we would expect that a fusion of Gata4 and VP16, which strongly activates transcription by recruiting pre-initiation complexes[16], would activate cardiac genes. Transfection of Gata4–VP16 alone did not activate *Actc1* (0 out of 7) (Fig. 3c) or *Nkx2-5* (0 out of 5), and Gata4–VP16 plus Baf60c induced *Actc1* efficiently (5 out of 5) but did not induce beating tissue (0 out of 9) (Fig. 3c). Thus, Baf60c must have functions separate from the recruitment of the pre-initiation complex. We then examined whether Baf60c is important for binding of Gata4 to target genes. We performed chromatin immunoprecipitation on cells isolated by FACS from transfected embryonic posterior mesoderm. Gata4 could not bind *Tnnt2* or *Nppa* regulatory regions in the absence of Baf60c, but inclusion of Baf60c resulted in detectable Gata4 occupancy on both genes

(Fig. 4c). The BAF complex ATPase Brg1 (also called Smarca4) was also detected only in the presence of Baf60c (Fig. 4c). Therefore, unlike the 'pioneer factor' function of Gata4 in liver development[17], Baf60c is required to allow Gata4 to bind these cardiac genes in non-cardiac mesoderm. Although the mechanism of this action is unknown, it may involve recognition of histone modifications by subunits of Baf60c-associated BAF complexes[6,7] or other interacting transcription factors[18], followed by remodelling of the chromatin at these loci or cooperative binding to target DNA.

Our results define a combination of three factors that can execute a cardiac transcriptional program and fully induce non-cardiac mesoderm to differentiate directly into contractile cardiac myocytes. The direct induction of ectopic beating tissue *in vivo* in amniotes is unprecedented; application of cardiogenic growth factors to chick embryonic mesoderm only induced early markers of cardiomyocyte lineage, such as *Nkx2-5* (refs 19, 20). In vertebrates with simpler hearts such as *Xenopus* or zebrafish, Mesp1 or Gata5 could induce cardiac differentiation, but only at low frequency (<10% transfected embryos)[10,21]. Mesp1 can promote cardiogenic mesoderm from embryonic stem cells[10–12], but was ineffective in the mouse, suggesting that it primarily promotes lineage specification rather than inducing

differentiation. We may have targeted the *in vivo* equivalent of epiblast stem cells (EpiSCs)[22,23] rather than mesoderm; this is unlikely as these should be rare or absent at stages at which we were able to induce cardiogenesis (E7.0–8.0), and loss of posterior mesoderm markers in induced cardiac cells supports targeting of mesoderm. It has been suggested that permissive cues from adjacent endoderm or ectoderm are important for cardiac differentiation; however, in chick and *Xenopus*, heart tissue could only be induced in explants[19,20,24,25], suggesting that strong negative cues prevent ectopic cardiac differentiation *in vivo*. Our results show that repressive influences are unlikely to be major components of cardiac program regulation or that they can be overcome by our combination of factors.

Overall, our data suggest that, in amniotes, an additional layer of tissue-specific gene regulation, in the form of Baf60c, has been superimposed on cardiac transcriptional regulators. Consistent with loss-of-function experiments[26], Gata4 was a key factor in initiating the cardiac program, whereas Tbx5 was required for full differentiation into contracting cardiomyocytes. Because Gata4 plus Baf60c induced *Nkx2-5*, the combined input from Gata4, Nkx2-5 and Tbx5, which synergize on target enhancers[1–3], provided sufficient information to activate the cardiac program. Unlike other contexts where DNA-binding factors were sufficient for transdifferentiation[27–29], Baf60c was required to potentiate the function of Gata4 and Tbx5, partly by allowing binding of Gata4 to cardiac loci. These findings reveal a novel mechanism for tissue-specific gene regulation by chromatin-remodelling complexes. Harnessing this potential in a permissive cell type might lead to the production of cardiac cells suitable for therapeutic use in diseased myocardium.

## METHODS SUMMARY

Mouse embryos were transfected with separate expression constructs for each factor and EGFP, and cultured as described[30] with modifications mainly in the timing and location of the transfection. Embryos were collected, fixed with 4% paraformaldehyde, and processed for whole-mount *in situ* hybridization, paraffin sections, or cryosections. Immunohistochemistry was performed on whole embryos, or on sections after antigen retrieval. Chromatin immunoprecipitation was performed on proteins cross-linked to DNA from cells FACS-sorted from 25–30 embryos for each condition.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

1.  Olson, E. N. Gene regulatory networks in the evolution and development of the heart. *Science* **313**, 1922–1927 (2006).
2.  Srivastava, D. Making or breaking the heart: from lineage determination to morphogenesis. *Cell* **126**, 1037–1048 (2006).
3.  Lickert, H. *et al.* Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature* **432**, 107–112 (2004).
4.  Simone, C. *et al.* p38 pathway targets SWI-SNF chromatin-remodeling complex to muscle-specific loci. *Nature Genet.* **36**, 738–743 (2004).
5.  Lessard, J. *et al.* An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* **55**, 201–215 (2007).
6.  Lange, M. *et al.* Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. *Genes Dev.* **22**, 2370–2384 (2008).
7.  Wu, J. I., Lessard, J. & Crabtree, G. R. Understanding the words of chromatin regulation. *Cell* **136**, 200–206 (2009).
8.  Creemers, E. E. *et al.* Coactivation of MEF2 by the SAP domain proteins myocardin and MASTR. *Mol. Cell* **23**, 83–96 (2006).
9.  Saga, Y. *et al.* MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. *Development* **126**, 3437–3447 (1999).
10. David, R. *et al.* MesP1 drives vertebrate cardiovascular differentiation through Dkk-1-mediated blockade of Wnt-signalling. *Nature Cell Biol.* **10**, 338–345 (2008).
11. Bondue, A. *et al.* Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. *Cell Stem Cell* **3**, 69–84 (2008).
12. Lindsley, R. C. *et al.* Mesp1 coordinately regulates cardiovascular fate restriction and epithelial-mesenchymal transition in differentiating ESCs. *Cell Stem Cell* **3**, 55–68 (2008).
13. Kinder, S. J. *et al.* The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. *Development* **126**, 4691–4701 (1999).
14. Pevny, L. *et al.* Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349**, 257–260 (1991).
15. Laugwitz, K. L. *et al.* Islet1 cardiovascular progenitors: a single source for heart lineages? *Development* **135**, 193–205 (2008).
16. Black, J. C. *et al.* A mechanism for coordinating chromatin modification and preinitiation complex assembly. *Mol. Cell* **23**, 809–818 (2006).
17. Cirillo, L. A. *et al.* Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
18. Lupien, M. *et al.* FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958–970 (2008).
19. Schultheiss, T. M., Burch, J. B. & Lassar, A. B. A role for bone morphogenetic proteins in the induction of cardiac myogenesis. *Genes Dev.* **11**, 451–462 (1997).
20. Marvin, M. J. *et al.* Inhibition of Wnt activity induces heart formation from posterior mesoderm. *Genes Dev.* **15**, 316–327 (2001).
21. Reiter, J. F. *et al.* Gata5 is required for the development of the heart and endoderm in zebrafish. *Genes Dev.* **13**, 2983–2995 (1999).
22. Brons, I. G. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191–195 (2007).
23. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–199 (2007).
24. Small, E. M. *et al.* Myocardin is sufficient and necessary for cardiac gene expression in *Xenopus*. *Development* **132**, 987–997 (2005).
25. Latinkic, B. V., Kotecha, S. & Mohun, T. J. Induction of cardiomyocytes by GATA4 in *Xenopus* ectodermal explants. *Development* **130**, 3865–3876 (2003).
26. Zhao, R. *et al.* Loss of both GATA4 and GATA6 blocks cardiac myocyte differentiation and results in acardia in mice. *Dev. Biol.* **317**, 614–619 (2008).
27. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987–1000 (1987).
28. Xie, H. *et al.* Stepwise reprogramming of B cells into macrophages. *Cell* **117**, 663–676 (2004).
29. Zhou, Q. *et al.* *In vivo* reprogramming of adult pancreatic exocrine cells to β-cells. *Nature* **455**, 627–632 (2008).
30. Yamamoto, M. *et al.* Nodal antagonists regulate formation of the anteroposterior axis of the mouse embryo. *Nature* **428**, 387–392 (2004).

**Author Contributions** J.K.T. performed all experimental work. J.K.T. and B.G.B. conceived the study. B.G.B. directed the study and wrote the paper.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.G.B. (bbruneau@gladstone.ucsf.edu) or J.K.T. (takeuchi.j.ab@m.titech.ac.jp).

## METHODS

**Mouse embryology.** Mouse embryos were transfected with separate expression constructs for each factor and EGFP, and cultured as described[30] with modifications mainly in the timing and location of the transfection. At embryonic day (E) 6.5–7.5, embryos were injected posteriorly under the visceral endoderm. Embryos were cultured for 36–48 h. It should be noted that there is variation between embryos as they are cultured, primarily reflecting slight differences in the stages of initial isolation and injection and in their individual development under culture conditions. Co-transfection of constructs was observed in at least 90% of transfected cells (Supplementary Fig. 2); the degree of EGFP signal does not always correlate with the level of induction of target genes, probably due to different transfection efficiencies for each vector in each cell. Expression constructs were all under the control of a CMV promoter.

**Immunohistochemistry, antibodies, histochemistry and *in situ* hybridization.** Embryos were collected, fixed with 4% paraformaldehyde and processed for paraffin sections or cryosections. Immunohistochemistry was performed on whole embryos or sections after antigen retrieval by microwave heating (for paraffin sections). Antisera used were anti-α-TM and anti-cTnT (both from the Developmental Studies Hybridoma Bank developed under the auspices of the NICHD and maintained by The University of Iowa, Department of Biological Sciences). Whole-mount *in situ* hybridization was performed with digoxigenin-labelled probes generated by *in vitro* transcription and standard procedures.

**Chromatin immunoprecipitation.** Proteins were cross-linked to DNA from cells FACS-sorted from 25–30 embryos by adding formaldehyde/PBS (final concentration 1%) to the culture medium. To stop cross-linking, glycine was added to a final concentration of 0.125 M. Cells were washed with PBS with protease inhibitors, and centrifuged. Cells were resuspended in 1% SDS, 10 mM EDTA and 50 mM Tris-HCl, pH 8.0, with protease inhibitor and sonicated. Samples were diluted (1/10 in 0.01% SDS, 1.1% Triton-X100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl, pH 8.0, and protease inhibitor), precleared with protein G beads and salmon sperm DNA, and centrifuged. Supernatants were incubated with primary antibody against the protein of interest for 4 h to over night at 4 °C. Antisera were anti-Brg1 (Upstate), anti-Gata4 (Santa Cruz C-20), or normal rabbit IgG (Santa Cruz, 2027). After multiple washes samples were eluted and cross-links reversed with 5 μl 5 M NaCl. Samples were purified before PCR amplification. Primer sequences are as follows: *Nppa* primers: 340F TGTTGCCAGGGAGAAAGAATCCT, 36R CAGCTCTTTGAGAAGGCA; *Tnnt2* primers: 50F ACCAACCTCCATCCGAAAGGT, 595R AGATGCCTTTGTTATT CCTGAAA.

# LETTERS

# Frequent inactivation of A20 in B-cell lymphomas

Motohiro Kato[1,2], Masashi Sanada[1,5], Itaru Kato[6], Yasuharu Sato[7], Junko Takita[1,2,3], Kengo Takeuchi[8], Akira Niwa[6], Yuyan Chen[1,2], Kumi Nakazaki[1,4,5], Junko Nomoto[9], Yoshitaka Asakura[9], Satsuki Muto[1], Azusa Tamura[1], Mitsuru Iio[1], Yoshiki Akatsuka[11], Yasuhide Hayashi[12], Hiraku Mori[13], Takashi Igarashi[2], Mineo Kurokawa[4], Shigeru Chiba[3], Shigeo Mori[14], Yuichi Ishikawa[8], Koji Okamoto[10], Kensei Tobinai[9], Hitoshi Nakagama[10], Tatsutoshi Nakahata[6], Tadashi Yoshino[7], Yukio Kobayashi[9] & Seishi Ogawa[1,5]

**A20 is a negative regulator of the NF-κB pathway and was initially identified as being rapidly induced after tumour-necrosis factor-α stimulation[1]. It has a pivotal role in regulation of the immune response and prevents excessive activation of NF-κB in response to a variety of external stimuli[2–7]; recent genetic studies have disclosed putative associations of polymorphic A20 (also called TNFAIP3) alleles with autoimmune disease risk[8,9]. However, the involvement of A20 in the development of human cancers is unknown. Here we show, using a genome-wide analysis of genetic lesions in 238 B-cell lymphomas, that A20 is a common genetic target in B-lineage lymphomas. A20 is frequently inactivated by somatic mutations and/or deletions in mucosa-associated tissue lymphoma (18 out of 87; 21.8%) and Hodgkin's lymphoma of nodular sclerosis histology (5 out of 15; 33.3%), and, to a lesser extent, in other B-lineage lymphomas. When re-expressed in a lymphoma-derived cell line with no functional A20 alleles, wild-type A20, but not mutant A20, resulted in suppression of cell growth and induction of apoptosis, accompanied by downregulation of NF-κB activation. The A20-deficient cells stably generated tumours in immunodeficient mice, whereas the tumorigenicity was effectively suppressed by re-expression of A20. In A20-deficient cells, suppression of both cell growth and NF-κB activity due to re-expression of A20 depended, at least partly, on cell-surface-receptor signalling, including the tumour-necrosis factor receptor. Considering the physiological function of A20 in the negative modulation of NF-κB activation induced by multiple upstream stimuli, our findings indicate that uncontrolled signalling of NF-κB caused by loss of A20 function is involved in the pathogenesis of subsets of B-lineage lymphomas.**

Malignant lymphomas of B-cell lineages are mature lymphoid neoplasms that arise from various lymphoid tissues[10,11]. To obtain a comprehensive registry of genetic lesions in B-lineage lymphomas, we performed a single nucleotide polymorphism (SNP) array analysis of 238 primary B-cell lymphoma specimens of different histologies, including 64 samples of diffuse large B-cell lymphomas (DLBCLs), 52 follicular lymphomas, 35 mantle cell lymphomas (MCLs), and 87 mucosa-associated tissue (MALT) lymphomas (Supplementary Table 1). Three Hodgkin's-lymphoma-derived cell lines were also analysed. Interrogating more than 250,000 SNP sites, this platform permitted the identification of copy number changes at an average resolution of less than 12 kilobases (kb). The use of large numbers of

SNP-specific probes is a unique feature of this platform, and combined with the CNAG/AsCNAR software, enabled accurate determination of 'allele-specific' copy numbers, and thus allowed for sensitive detection of loss of heterozygosity (LOH) even without apparent copy-number reduction, in the presence of up to 70–80% normal cell contamination[12,13].

Lymphoma genomes underwent a wide range of genetic changes, including numerical chromosomal abnormalities and segmental gains and losses of chromosomal material (Supplementary Fig. 1), as well as copy-number-neutral LOH, or uniparental disomy (Supplementary Fig. 2). Each histology type had a unique genomic signature, indicating a distinctive underlying molecular pathogenesis for different histology types (Fig. 1a and Supplementary Fig. 3). On the basis of the genomic signatures, the initial pathological diagnosis of MCL was re-evaluated and corrected to DLBCL in two cases. Although most copy number changes involved large chromosomal segments, a number of regions showed focal gains and deletions, accelerating identification of their candidate gene targets. After excluding known copy number variations, we identified 46 loci showing focal gains (19 loci) or deletions (27 loci) (Supplementary Tables 2 and 3 and Supplementary Fig. 4).

Genetic lesions on the NF-κB pathway were common in B-cell lymphomas and found in approximately 40% of the cases (Supplementary Table 1), underpinning the importance of aberrant NF-κB activation in lymphomagenesis[11,14] in a genome-wide fashion. They included focal gain/amplification at the REL locus (16.4%) (Fig. 1b) and TRAF6 locus (5.9%), as well as focal deletions at the PTEN locus (5.5%) (Supplementary Figs 1 and 4). However, the most striking finding was the common deletion at 6q23.3 involving a 143-kb segment. It exclusively contained the A20 gene (also called TNFAIP3), a negative regulator of NF-κB activation[3–7,15] (Fig. 1b), which was previously reported as a candidate target of 6q23 deletions in ocular lymphoma[16]. LOH involving the A20 locus was found in 50 cases, of which 12 showed homozygous deletions as determined by the loss of both alleles in an allele-specific copy number analysis (Fig. 1b, Table 1 and Supplementary Table 4). On the basis of this finding, we searched for possible tumour-specific mutations of A20 by genomic DNA sequencing of entire coding exons of the gene in the same series of lymphoma samples (Supplementary Fig. 5). Because two out of the three Hodgkin's-lymphoma-derived cell lines had biallelic A20 deletions/mutations (Supplementary Fig. 6), 24 primary samples from Hodgkin's lymphoma were also analysed for mutations, where

**Figure 1 | Genomic signatures of different B-cell lymphomas and common genetic lesions at 2p16-15 and 6q23.3 involving NF-κB pathway genes.**
**a**, Twenty-nine genetic lesions were found in more than 10% in at least one histology and used for clustering four distinct histology types of B-lineage lymphomas. The frequency of each genetic lesion in each histology type is colour-coded. FL, follicular lymphoma; UPD, uniparental disomy.
**b**, Recurrent genetic changes are depicted based on CNAG output of the SNP array analysis of 238 B-lineage lymphoma samples, which include gains at the *REL* locus on 2p16-15 (left panel) and the *A20* locus on 6q23.3 (right

panel). Regions showing copy number gain or loss are indicated by horizontal lines. Four histology types are indicated by different colours, where high-grade amplifications and homozygous deletions are shown by darker shades to discriminate from simple gains (copy number ≤4) and losses (copy number = 1) (lighter shades). **c**, Point mutations and small nucleotide insertions and deletions in the *A20* (*TNFAIP3*) gene caused premature truncation of A20 in most cases. Altered amino acids caused by frame shifts are indicated by green bars.

genomic DNA was extracted from 150 microdissected CD30-positive tumour cells (Reed–Sternberg cells) for each sample. *A20* mutations were found in 18 out of 265 lymphoma samples (6.8%) (Table 1), among which 13 mutations, including nonsense mutations (3 cases), frame-shift insertions/deletions (9 cases), and a splicing donor site mutation (1 case) were thought to result in premature termination of translation (Fig. 1c). Four missense mutations and one intronic mutation were identified in five microdissected Hodgkin's lymphoma samples. They were not found in the surrounding normal tissues, and thus, were considered as tumour-specific somatic changes.

In total, biallelic *A20* lesions were found in 31 out of 265 lymphoma samples including 3 Hodgkin's lymphoma cell lines. Quantitative analysis of SNP array data suggested that these *A20* lesions were present in the major tumour fraction within the samples (Supplementary Fig. 7). Inactivation of *A20* was most frequent in MALT lymphoma (18 out of 87) and Hodgkin's lymphoma (7 out of 27), although it was also found in DLBCL (5 out of 64) and follicular lymphoma (1 out of 52) at lower frequencies. In MALT lymphoma, biallelic *A20* lesions were confirmed in 18 out of 24 cases (75.0%) with LOH involving the 6q23.3 segment (Supplementary Fig. 8). Considering the limitation in detecting very small homozygous deletions, *A20* was thought to be the target of 6q23 LOH in MALT lymphoma. On the other hand, the 6q23 LOHs in other histology types tended to be extended into more centromeric regions and less frequently accompanied biallelic *A20* lesions (Supplementary Fig. 8 and Supplementary Table 4), indicating that they might be more

heterogeneous with regard to their gene targets. We were unable to analyse Hodgkin's lymphoma samples using SNP arrays owing to insufficient genomic DNA obtained from microdissected samples, and were likely to underestimate the frequency of *A20* inactivation in Hodgkin's lymphoma because we might fail to detect a substantial proportion of cases with homozygous deletions, which explained 50% (12 out of 24) of *A20* inactivation in other histology types. *A20* mutations in Hodgkin's lymphoma were exclusively found in nodular sclerosis classical Hodgkin's lymphoma (5 out of 15) but not in other histology types (0 out of 9), although the possible association requires further confirmation in additional cases.

A20 is a key regulator of NF-κB signalling, negatively modulating NF-κB activation through a wide variety of cell surface receptors and viral proteins, including tumour-necrosis factor (TNF) receptors, toll-like receptors, CD40, as well as Epstein–Barr-virus-associated LMP1 protein[2,5,17,18]. To investigate the role of A20 inactivation in lymphomagenesis, we re-expressed wild-type A20 under a *Tet*-inducible promoter in a lymphoma-derived cell line (KM-H2) that had no functional *A20* alleles (Supplementary Fig. 6), and examined the effect of A20 re-expression on cell proliferation, survival and downstream NF-κB signalling pathways. As shown in Fig. 2a–c and Supplementary Fig. 9, re-expression of wild-type A20 resulted in the suppression of cell proliferation and enhanced apoptosis, and in the concomitant accumulation of IκBβ and IκBε, and downregulation of NF-κB activity. In contrast, re-expression of two lymphoma-derived A20 mutants, A20[532Stop] or A20[750Stop], failed to show growth suppression, induction of apoptosis, accumulation of IκBβ and IκBε or downregulation of

**Table 1 | Inactivation of A20 in B-lineage lymphomas**

| Histology | Tissue | Sample | Allele | Uniparental disomy | Exon | Mutation | Biallelic inactivation |
|---|---|---|---|---|---|---|---|
| DLBCL | | | | | | | 5 out of 64 (7.8%) |
| | Lymph node | DLBCL008 | −/− | No | — | — | |
| | Lymph node | DLBCL016 | +/− | No | Ex2 | 329insA | |
| | Lymph node | DLBCL022 | −/− | No | — | — | |
| | Lymph node | DLBCL028 | −/− | Yes | — | — | |
| | Lymph node | MCL008* | −/− | Yes | — | — | |
| Follicular lymphoma | | | | | | | 1 out of 52 (1.9%) |
| | Lymph node | FL024 | −/− | No | — | — | |
| MCL | | | | | | | 0 out of 35 (0%) |
| MALT | | | | | | | 18 out of 87 (21.8%) |
| Stomach | | | | | | | 3 out of 23 (13.0%) |
| | Gastric mucosa | MALT013 | +/+ | Yes | Ex5 | 705insG | |
| | Gastric mucosa | MALT014 | +/+ | Yes | Ex3 | Ex3 donor site>A | |
| | Gastric mucosa | MALT036 | +/− | No | Ex7 | delintron6−Ex7† | |
| Eye | | | | | | | 13 out of 43 (30.2%) |
| | Ocular adnexa | MALT008 | −/− | No | — | — | |
| | Ocular adnexa | MALT017 | −/− | No | — | — | |
| | Ocular adnexa | MALT051 | +/− | No | Ex7 | 1943delTG | |
| | Ocular adnexa | MALT053 | +/+ | Yes | Ex6 | 1016G>A(stop) | |
| | Ocular adnexa | MALT054 | +/− | No | Ex3 | 502delTC | |
| | Ocular adnexa | MALT055 | −/− | No | — | — | |
| | Ocular adnexa | MALT066 | +/− | No | Ex7 | 1581insA | |
| | Ocular adnexa | MALT067 | −/− | No | — | — | |
| | Ocular adnexa | MALT082 | −/− | Yes | — | — | |
| | Ocular adnexa | MALT084 | −/− | Yes | — | — | |
| | Ocular adnexa | MALT085 | +/+ | Yes | Ex7 | 1435insG | |
| | Ocular adnexa | MALT086 | +/+ | Yes | Ex6 | 878C>T(stop) | |
| | Ocular adnexa | MALT087 | +/+ | Yes | Ex9 | 2304delGG | |
| Lung | | | | | | | 2 out of 12 (16.7%) |
| | Lung | MALT042 | −/− | No | — | — | |
| | Lung | MALT047 | +/+ | Yes | Ex9 | 2281insT | |
| Other‡ | | | | | | | 0 out of 9 (0%) |
| Hodgkin's lymphoma | | | | | | | 7 out of 27 (26.0%) |
| NSHL | Lymph node | HL10 | ND | ND | Ex7 | 1777G>A(V571I) | |
| NSHL | Lymph node | HL12 | ND | ND | Ex7 | 1156A>G(R364G) | |
| NSHL | Lymph node | HL21 | ND | ND | Ex4 | 569G>A(stop) | |
| NSHL | Lymph node | HL24 | ND | ND | Ex3 | 1487C>A(T474N) | |
| NSHL | Lymph node | HL23 | ND | ND | — | Intron 3§ | |
| | Cell line | KM-H2 | −/− | No | — | | |
| | Cell line | HDLM2 | +/− | No | Ex4 | 616ins29bp‖ | |
| Total | | | | | | | 31 out of 265 (11.7%) |

DLBCL, diffuse large B-cell lymphoma; MALT, MALT lymphoma; MCL, mantle cell lymphoma; ND, not determined because SNP array analysis was not performed; NSHL, nodular sclerosis classical Hodgkin's lymphoma.
\* Diagnosis was changed based on the genomic data, which was confirmed by re-examination of pathology.
† Deletion including the boundary of intron 6 and exon 7 (see also Supplementary Fig. 5b).
‡ Including 1 parotid gland, 1 salivary gland, 2 colon and 5 thyroid cases.
§ Insertion of CTC at −19 bases from the beginning of exon 3.
‖ Insertion of TGGCTTCCACAGACACACCCATGGCCCGA.

NF-κB activity (Fig. 2a–c), indicating that these were actually loss-of-function mutations. To investigate the role of A20 inactivation in lymphomagenesis *in vivo*, *A20*- and mock-transduced KM-H2 cells were transplanted in NOD/SCID/γ$_c$$^{null}$ (NOG) mice[19], and their tumour formation status was examined for 5 weeks with or without induction of wild-type A20 by tetracycline administration. As shown in Fig. 2d, mock-transduced cells developed tumours at the injected sites, whereas the *Tet*-inducible *A20*-transduced cells generated tumours only in the absence of A20 induction (Supplementary Table 5), further supporting the tumour suppressor role of A20 in lymphoma development.

Given the mode of negative regulation of NF-κB signalling, we next investigated the origins of NF-κB activity that was deregulated by A20 loss in KM-H2 cells. The conditioned medium prepared from a 48-h serum-free KM-H2 culture had increased NF-κB upregulatory activity compared with fresh serum-free medium, which was inhibited by re-expression of A20 (Fig. 3a). KM-H2 cells secreted two known ligands for TNF receptor—TNF-α and lymphotoxin-α (Supplementary Fig. 10)[20]—and adding neutralizing antibodies against these cytokines into cultures significantly suppressed their cell growth and NF-κB activity without affecting the levels of their overall suppression after A20

induction (Fig. 3b, d). In addition, recombinant TNF-α and/or lymphotoxin-α added to fresh serum-free medium promoted cell growth and NF-κB activation in KM-H2 culture, which were again suppressed by re-expression of A20 (Fig. 3c, e). Although our data in Fig. 3 also show the presence of factors other than TNF-α and lymphotoxin-α in the KM-H2-conditioned medium—as well as some intrinsic pathways in the cell (Fig. 3a)—that were responsible for the A20-dependent NF-κB activation, these results indicate that both cell growth and NF-κB activity that were upregulated by A20 inactivation depend at least partly on the upstream stimuli that evoked the NF-κB-activating signals.

Aberrant activation of the NF-κB pathway is a hallmark of several subtypes of B-lineage lymphomas, including Hodgkin's lymphoma, MALT lymphoma, and a subset of DLBCL, as well as other lymphoid neoplasms[11,14], where a number of genetic alterations of NF-κB signalling pathway genes[21–25], as well as some viral proteins[26,27], have been implicated in the aberrant activation of the NF-κB pathway[14]. Thus, frequent inactivation of A20 in Hodgkin's lymphoma and MALT and other lymphomas provides a novel insight into the molecular pathogenesis of these subtypes of B-lineage lymphomas through deregulated NF-κB activation. Because A20 provides a
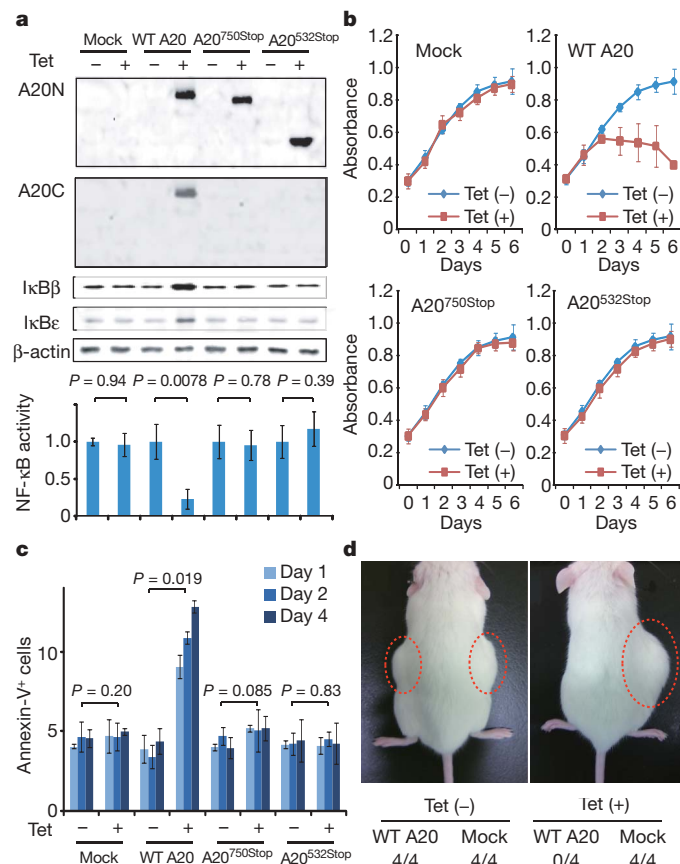
**Figure 2 | Effects of wild-type and mutant A20 re-expressed in a lymphoma cell line that lacks the normal *A20* gene. a**, Western blot analyses of wild-type (WT) and mutant (A20^532Stop and A20^750Stop) A20, as well as IκBβ and IκBε, in KM-H2 cells, in the presence or absence of tetracycline treatment (top panels). A20N and A20C are polyclonal antisera raised against N-terminal and C-terminal A20 peptides, respectively. β-actin blots are provided as a control. NF-κB activities are expressed as mean absorbance ± s.d. (*n* = 6) in luciferase assays (bottom panel).
**b**, Proliferation of KM-H2 cells stably transduced with plasmids for mock and *Tet*-inducible wild-type A20, A20^532Stop and A20^750Stop was measured using a cell counting kit in the presence (red lines) or absence (blue lines) of tetracycline. Mean absorbance ± s.d. (*n* = 5) is plotted. **c**, The fractions of Annexin-V-positive KM-H2 cells transduced with various *Tet*-inducible A20 constructs were measured by flow cytometry after tetracycline treatment and the mean values (±s.d., *n* = 3) are plotted. **d**, *In vivo* tumorigenicity was assayed by inoculating 7 × 10^6 KM-H2 cells transduced with mock or *Tet*-inducible wild-type A20 in NOG mice, with (right panel) or without (left panel) tetracycline administration.

negative feedback mechanism in the regulation of NF-κB signalling pathways upon a variety of stimuli, aberrant activation of NF-κB will be a logical consequence of A20 inactivation. However, there is also the possibility that the aberrant NF-κB activity of A20-inactivated lymphoma cells is derived from upstream stimuli, which may be from the cellular environment. In this context, it is intriguing that MALT lymphoma usually arises at the site of chronic inflammation caused by infection or autoimmune disorders and may show spontaneous regression after eradication of infectious organisms[28]; furthermore, Hodgkin's lymphoma frequently shows deregulated cytokine production from Reed–Sternberg cells and/or surrounding reactive cells[29]. Detailed characterization of the NF-κB pathway regulated by A20 in both normal and neoplastic B lymphocytes will promote our understanding of the precise roles of A20 inactivation in the pathogenesis of these lymphoma types. Our finding underscores the importance of genome-wide approaches in the identification of genetic targets in human cancers.
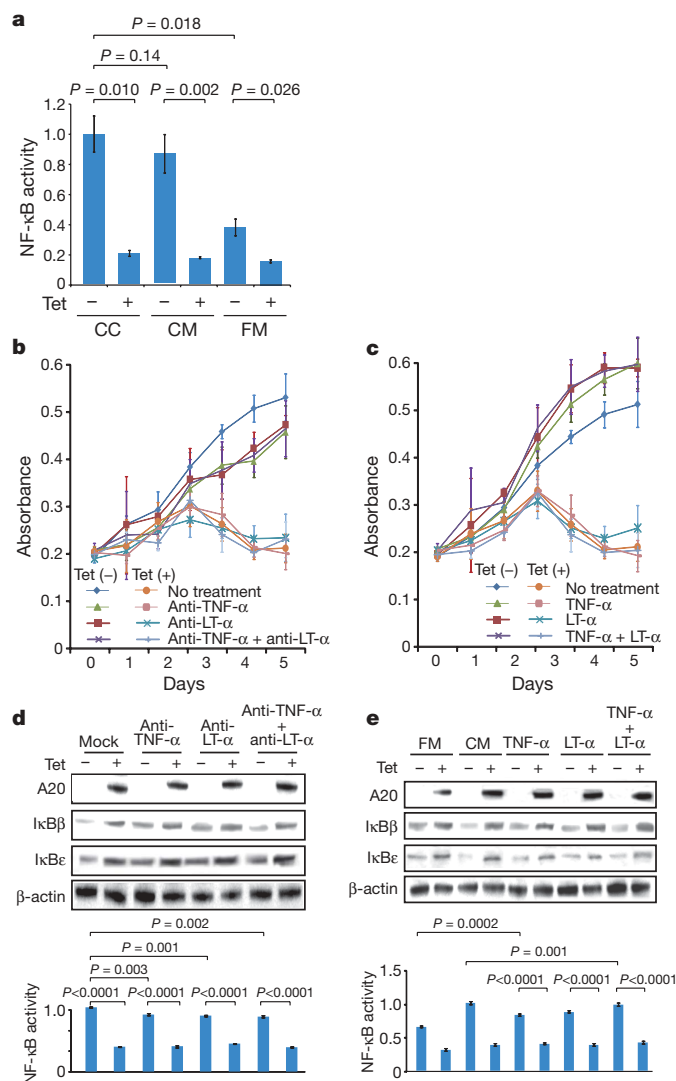
**Figure 3 | Tumour suppressor role of A20 under external stimuli. a**, NF-κB activity in KM-H2 cells was measured 30 min after cells were inoculated into fresh medium (FM) or KM-H2-conditioned medium (CM) obtained from the 48-h culture of KM-H2, and was compared with the activity after 48 h continuous culture of KM-H2 (CC). A20 was induced 12 h before inoculation in Tet (+) groups. **b, c**, Effects of neutralizing antibodies against TNF-α and lymphotoxin-α (LTα) (**b**) and of recombinant TNF-α and LT-α added to the culture (**c**) on cell growth were evaluated in the presence (Tet (+)) or absence (Tet (−)) of A20 induction. Cell numbers were measured using a cell counting kit and are plotted as their mean absorbance ± s.d. (*n* = 6). **d, e**, Effects of the neutralizing antibodies (**d**) and the recombinant cytokines added to the culture (**e**) on NF-κB activities and the levels of IκBβ and IκBε after 48 h culture with (Tet (+)) or without (Tet (−)) tetracycline treatment. NF-κB activities are expressed as mean absorbance ± s.d. (*n* = 6) in luciferase assays.

**METHODS SUMMARY**

Genomic DNA from 238 patients with non-Hodgkin's lymphoma and three Hodgkin's-lymphoma-derived cell lines was analysed using GeneChip SNP genotyping microarrays (Affymetrix). This study was approved by the ethics boards of the University of Tokyo, National Cancer Institute Hospital, Okayama University, and the Cancer Institute of the Japanese Foundation of Cancer Research. After appropriate normalization of mean array intensities, signal ratios between tumours and anonymous normal references were calculated in an allele-specific manner, and allele-specific copy numbers were inferred from the observed signal ratios based on the hidden Markov model using CNAG/AsCNAR software (http://www.genome.umin.jp). *A20* mutations were examined by directly sequencing genomic DNA using a set of primers (Supplementary Table 6). Full-length cDNAs of wild-type and mutant *A20* were introduced into a

lentivirus vector, pLenti4/TO/V5-DEST (Invitrogen), with a *Tet*-inducible promoter. Viral stocks were prepared by transfecting the vector plasmids into 293FT cells (Invitrogen) using the calcium phosphate method and then infected to the KM-H2 cell line. Proliferation of KM-H2 cells was measured using a Cell Counting Kit (Dojindo). Western blot analyses and luciferase assays were performed as previously described. NF-κB activity was measured by luciferase assays in KM-H2 cells stably transduced with a reporter plasmid having an NF-κB response element, pGL4.32 (Promega). Apoptosis of KM-H2 upon A20 induction was evaluated by counting Annexin-V-positive cells by flow cytometry. For *in vivo* tumorigenicity assays, $7 \times 10^6$ KM-H2 cells were transduced with the *Tet*-inducible *A20* gene and those with a mock vector were inoculated on the contralateral sides in eight NOG mice[19] and examined for their tumour formation with ($n = 4$) or without ($n = 4$) tetracycline administration. Full copy number data of the 238 lymphoma samples will be accessible from the Gene Expression Omnibus (GEO, http://ncbi.nlm.nih.gov/geo/) with the accession number GSE12906.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

1.  Dixit, V. M. *et al.* Tumor necrosis factor-α induction of novel gene products in human endothelial cells including a macrophage-specific chemotaxin. *J. Biol. Chem.* **265**, 2973–2978 (1990).
2.  Song, H. Y., Rothe, M. & Goeddel, D. V. The tumor necrosis factor-inducible zinc finger protein A20 interacts with TRAF1/TRAF2 and inhibits NF-κB activation. *Proc. Natl Acad. Sci. USA* **93**, 6721–6725 (1996).
3.  Lee, E. G. *et al.* Failure to regulate TNF-induced NF-κB and cell death responses in A20-deficient mice. *Science* **289**, 2350–2354 (2000).
4.  Boone, D. L. *et al.* The ubiquitin-modifying enzyme A20 is required for termination of Toll-like receptor responses. *Nature Immunol.* **5**, 1052–1060 (2004).
5.  Wang, Y. Y., Li, L., Han, K. J., Zhai, Z. & Shu, H. B. A20 is a potent inhibitor of TLR3- and Sendai virus-induced activation of NF-κB and ISRE and IFN-β promoter. *FEBS Lett.* **576**, 86–90 (2004).
6.  Wertz, I. E. *et al.* De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF-κB signalling. *Nature* **430**, 694–699 (2004).
7.  Heyninck, K. & Beyaert, R. A20 inhibits NF-κB activation by dual ubiquitin-editing functions. *Trends Biochem. Sci.* **30**, 1–4 (2005).
8.  Graham, R. R. *et al.* Genetic variants near *TNFAIP3* on 6q23 are associated with systemic lupus erythematosus. *Nature Genet.* **40**, 1059–1061 (2008).
9.  Musone, S. L. *et al.* Multiple polymorphisms in the *TNFAIP3* region are independently associated with systemic lupus erythematosus. *Nature Genet.* **40**, 1062–1064 (2008).
10. Jaffe, E. S., Harris, N. L., Stein, H. & Vardiman, J. W. *World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of Hematopoietic and Lymphoid Tissues* (IARC Press, 2001).
11. Klein, U. & Dalla-Favera, R. Germinal centres: role in B-cell physiology and malignancy. *Nature Rev. Immunol.* **8**, 22–33 (2008).
12. Nannya, Y. *et al.* A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res.* **65**, 6071–6079 (2005).
13. Yamamoto, G. *et al.* Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am. J. Hum. Genet.* **81**, 114–126 (2007).
14. Jost, P. J. & Ruland, J. Aberrant NF-κB signaling in lymphoma: mechanisms, consequences, and therapeutic implications. *Blood* **109**, 2700–2707 (2007).
15. Durkop, H., Hirsch, B., Hahn, C., Foss, H. D. & Stein, H. Differential expression and function of A20 and TRAF1 in Hodgkin lymphoma and anaplastic large cell lymphoma and their induction by CD30 stimulation. *J. Pathol.* **200**, 229–239 (2003).
16. Honma, K. *et al. TNFAIP3* is the target gene of chromosome band 6q23.3-q24.1 loss in ocular adnexal marginal zone B cell lymphoma. *Genes Chromosom. Cancer* **47**, 1–7 (2008).
17. Sarma, V. *et al.* Activation of the B-cell surface receptor CD40 induces A20, a novel zinc finger protein that inhibits apoptosis. *J. Biol. Chem.* **270**, 12343–12346 (1995).
18. Fries, K. L., Miller, W. E. & Raab-Traub, N. The A20 protein interacts with the Epstein-Barr virus latent membrane protein 1 (LMP1) and alters the LMP1/ TRAF1/TRADD complex. *Virology* **264**, 159–166 (1999).
19. Hiramatsu, H. *et al.* Complete reconstitution of human lymphocytes from cord blood CD34+ cells using the NOD/SCID/γ^null mice model. *Blood* **102**, 873–880 (2003).
20. Hsu, P. L. & Hsu, S. M. Production of tumor necrosis factor-α and lymphotoxin by cells of Hodgkin's neoplastic cell lines HDLM-1 and KM-H2. *Am. J. Pathol.* **135**, 735–745 (1989).
21. Dierlamm, J. *et al.* The apoptosis inhibitor gene *API2* and a novel 18q gene, *MLT*, are recurrently rearranged in the t(11;18)(q21;q21) associated with mucosa-associated lymphoid tissue lymphomas. *Blood* **93**, 3601–3609 (1999).
22. Willis, T. G. *et al.* Bcl10 is involved in t(1;14)(p22;q32) of MALT B cell lymphoma and mutated in multiple tumor types. *Cell* **96**, 35–45 (1999).
23. Joos, S. *et al.* Classical Hodgkin lymphoma is characterized by recurrent copy number gains of the short arm of chromosome 2. *Blood* **99**, 1381–1387 (2002).
24. Martin-Subero, J. I. *et al.* Recurrent involvement of the *REL* and *BCL11A* loci in classical Hodgkin lymphoma. *Blood* **99**, 1474–1477 (2002).
25. Lenz, G. *et al.* Oncogenic *CARD11* mutations in human diffuse large B cell lymphoma. *Science* **319**, 1676–1679 (2008).
26. Deacon, E. M. *et al.* Epstein-Barr virus and Hodgkin's disease: transcriptional analysis of virus latency in the malignant cells. *J. Exp. Med.* **177**, 339–349 (1993).
27. Yin, M. J. *et al.* HTLV-I Tax protein binds to MEKK1 to stimulate IκB kinase activity and NF-κB activation. *Cell* **93**, 875–884 (1998).
28. Isaacson, P. G. & Du, M. Q. MALT lymphoma: from morphology to molecules. *Nature Rev. Cancer* **4**, 644–653 (2004).
29. Skinnider, B. F. & Mak, T. W. The role of cytokines in classical Hodgkin lymphoma. *Blood* **99**, 4283–4297 (2002).

**Author Contributions** M.Ka., K.N. and M.S. performed microarray experiments and subsequent data analyses. M.Ka., Y.C., K.Ta., J.T., J.N., M.I., A.T. and Y.K. performed mutation analysis of *A20*. M.Ka., S.Mu., M.S., Y.C. and Y.Ak. conducted functional assays of mutant A20. Y.S., K.Ta., Y.As., H.M., M.Ku., S.Mo., S.C., Y.K., K.To. and Y.I. prepared tumour specimens. I.K., K.O., A.N., H.N. and T.N. conducted *in vivo* tumorigenicity experiments in NOG/SCID mice. T.I., Y.H., T.Y., Y.K. and S.O. designed overall studies, and S.O. wrote the manuscript. All authors discussed the results and commented on the manuscript.

**Author Information** The copy number data as well as the raw microarray data will be accessible from the GEO (http://ncbi.nlm.nih.gov/geo/) with the accession number GSE12906. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.O. (sogawa-tky@umin.ac.jp) or Y.K. (ykkobaya@ncc.go.jp).

# METHODS

**Specimens.** Primary tumour specimens were obtained from patients who were diagnosed with DLBCL, follicular lymphoma, MCL, MALT lymphoma, or classical Hodgkin's lymphoma. In total, 238 primary lymphoma specimens listed in Supplementary Table 1 were subjected to SNP array analysis. Three Hodgkin's-lymphoma-derived cell lines (KM-H2, HDLM2, L540) were obtained from Hayashibara Biochemical Laboratories, Inc., Fujisaki Cell Center and were also analysed by SNP array analysis.

**Microarray analysis.** High-molecular-mass DNA was isolated from tumour specimens and subjected to SNP array analysis using GeneChip Mapping 50K and/or 250K arrays (Affymetrix). The scanned array images were processed with Gene Chip Operation software (GCOS), followed by SNP calls using GTYPE. Genome-wide copy number measurements and LOH detection were performed using CNAG/AsCNAR software[12,13].

**Mutation analysis.** Mutations in the *A20* gene were examined in 265 samples of B-lineage lymphoma, including 62 DLBCLs, 52 follicular lymphomas, 87 MALTs, 37 MCLs and 3 Hodgkin's-lymphoma-derived cell lines and 24 primary Hodgkin's lymphoma samples, by direct sequencing using an ABI PRISM 3130xl Genetic Analyser (Applied Biosystems). To analyse primary Hodgkin's lymphoma samples in which CD30-positive tumour cells (Reed–Sternberg cells) account for only a fraction of the specimen, 150 Reed–Sternberg cells were collected for each 10 μm slice of a formalin-fixed block immunostained for CD30 by laser-capture microdissection (ASLMD6000, Leica), followed by genomic DNA extraction using QIAamp DNA Micro kit (Qiagen). The primer sets used in this study are listed in Supplementary Table 6.

**Functional analysis of wild-type and mutant A20.** Full-length cDNA for wild-type *A20* was isolated from total RNA extracted from an acute myeloid leukaemia-derived cell line, CTS, and subcloned into a lentivirus vector (pLenti4/TO/V5-DEST, Invitrogen). cDNAs for mutant A20 were generated by PCR amplification using mutagenic primers (Supplementary Table 6), and introduced into the same lentivirus vector. Forty-eight hours after transfection of each plasmid into 293FT cells using the calcium phosphate method, lentivirus stocks were obtained from ultrafiltration using Amicon Ultra (Millipore), and used to infect KM-H2 cells to generate stable transfectants of mock, wild-type and mutant A20. Each KM-H2 derivative cell line was further transduced stably with a reporter plasmid (pGL4.32, Promega) containing a luciferase gene under an NF-κB-responsive element by electroporation using Nucleofector reagents (Amaxa).

**Assays for cell proliferation and NF-κB activity.** Proliferation of the KM-H2 derivative cell lines was assayed in triplicate using a Cell Counting Kit (Dojindo). The mean absorption of five independent assays was plotted with s.d. for each derivative line. Two independent KM-H2-derived cell lines were used for each experiment. The NF-κB activity in KM-H2 derivatives for A20 mutants was evaluated by luciferase assays using a PiccaGene Luciferase Assay Kit (TOYO B-Net Co.). Each assay was performed in triplicate and the mean absorption of five independent experiments was plotted with s.d.

**Western blot analyses.** Polyclonal anti-sera against N-terminal (anti-A20N) and C-terminal (anti-A20C) A20 peptides were generated by immunizing rabbits with these peptides (LSNMRKAVKIRERTPEDIC for anti-A20N and CFQFKQMYG for anti-A20C, respectively). Total cell lysates from KM-H2 cells were separated on 7.5% polyacrylamide gel and subjected to western blot analysis using antibodies to A20 (anti-A20N and anti-A20C), IκBα (sc-847), IκBβ (sc-945), IκBγ (sc-7155) and actin (sc-8432) (Santa Cruz Biotechnology).

**Functional analyses of wild-type and mutant A20.** Each KM-H2 derivative cell line stably transduced with various *Tet*-inducible A20 constructs was cultured in serum-free medium in the presence or absence of A20 induction using 1 μg ml$^{-1}$ of tetracycline, and cell number was counted every day. $1 \times 10^6$ cells of each KM-H2 derivative cell line was analysed for their intracellular levels of IκBβ and IκBε and for NF-κB activities by western blot analyses and luciferase assays, respectively, 12 h after the beginning of cell culture. Effects of human recombinant TNF-α and lymphotoxin-α (210-TA and 211-TB, respectively, R&D Systems) on the NF-κB pathway and cell proliferation were evaluated by adding both cytokines into 10 ml of serum-free cell culture at a concentration of 200 pg ml$^{-1}$. For cell proliferation assays, culture medium was half replaced every 12 h to minimize the side-effects of autocrine cytokines. Intracellular levels of IκBβ, IκBε and NF-κB were examined 12 h after the beginning of the cell culture. To evaluate the effect of neutralizing TNF-α and lymphotoxin-α, $1 \times 10^6$ of KM-H2 cells transduced with both *Tet*-inducible A20 and the NF-κB-luciferase reporter in serum-free media for 36 h, and thereafter neutralizing antibodies against TNF-α (MAB210, R&D Systems) and/or lymphotoxin-α (AF-211-NA, R&D Systems) were added to the media at a concentration of 200 pg ml$^{-1}$. After the extended culture during 12 h with or without 1 μg ml$^{-1}$ tetracycline, the intracellular levels of IκBβ and IκBε and NF-κB activities were examined by western blot analysis and luciferase assays, respectively. To examine the effects of A20 re-expression on apoptosis, $1 \times 10^6$ KM-H2 cells were cultured for 4 days in 10 ml medium with or without *Tet* induction. After staining with phycoerythrin-conjugated anti-Annexin-V (ID556422, Becton Dickinson), Annexin-V-positive cells were counted by flow cytometry at the indicated times.

**_In vivo_ tumorigenicity assays.** KM-H2 cells transduced with a mock or *Tet*-inducible wild-type *A20* gene were inoculated into NOG mice and their tumorigenicity was examined for 5 weeks with or without tetracycline administration. Injections of $7 \times 10^6$ cells of each KM-H2 cell line were administered to two opposite sites in four mice. Tetracycline was administered in drinking water at a concentration of 200 μg ml$^{-1}$.

**ELISA.** Concentrations of TNF-α, lymphotoxin-α, IL-1, IL-2, IL-4, IL-6, IL-12, IL-18 and TGF-β in the culture medium were measured after 48 h using ELISA. For those cytokines detectable after 48-h culture (TNFα, LTα, and IL-6), their time course was examined further using the Quantikine ELISA kit (R&D Systems).

**Statistical analysis.** Significance of the difference in NF-κB activity between two given groups was evaluated using a paired *t*-test, in which the data from each independent luciferase assay were paired to calculate test statistics. To evaluate the effect of A20 re-expression in KM-H2 cells on apoptosis, the difference in the fractions of Annexin-V-positive cells between Tet (+) and Tet (−) groups was also tested by a paired *t*-test for assays, in which the data from the assays performed on the same day were paired.

# LETTERS

# Mutations of multiple genes cause deregulation of NF-κB in diffuse large B-cell lymphoma

Mara Compagno[1], Wei Keat Lim[2], Adina Grunn[1], Subhadra V. Nandula[1,3], Manisha Brahmachary[1], Qiong Shen[1], Francesco Bertoni[4], Maurilio Ponzoni[5], Marta Scandurra[4], Andrea Califano[1,2], Govind Bhagat[1,3], Amy Chadburn[6], Riccardo Dalla-Favera[1,3,7] & Laura Pasqualicci[1,3]

**Diffuse large B-cell lymphoma (DLBCL), the most common form of lymphoma in adulthood, comprises multiple biologically and clinically distinct subtypes including germinal centre B-cell-like (GCB) and activated B-cell-like (ABC) DLBCL[1]. Gene expression profile studies have shown that its most aggressive subtype, ABC-DLBCL, is associated with constitutive activation of the NF-κB transcription complex[2]. However, except for a small fraction of cases[3], it remains unclear whether NF-κB activation in these tumours represents an intrinsic program of the tumour cell of origin or a pathogenetic event. Here we show that >50% of ABC-DLBCL and a smaller fraction of GCB-DLBCL carry somatic mutations in multiple genes, including negative (*TNFAIP3*, also called *A20*) and positive (*CARD11*, *TRAF2*, *TRAF5*, *MAP3K7* (*TAK1*) and *TNFRSF11A* (*RANK*)) regulators of NF-κB. Of these, the *A20* gene, which encodes a ubiquitin-modifying enzyme involved in termination of NF-κB responses, is most commonly affected, with ~30% of patients displaying biallelic inactivation by mutations and/or deletions. When reintroduced in cell lines carrying biallelic inactivation of the gene, A20 induced apoptosis and cell growth arrest, indicating a tumour suppressor role. Less frequently, missense mutations of *TRAF2* and *CARD11* produce molecules with significantly enhanced ability to activate NF-κB. Thus, our results demonstrate that NF-κB activation in DLBCL is caused by genetic lesions affecting multiple genes, the loss or activation of which may promote lymphomagenesis by leading to abnormally prolonged NF-κB responses.**

DLBCL represents a heterogeneous disease in terms of genetic, phenotypic and clinical features. Accordingly, genome-wide expression profile studies revealed the existence of several DLBCL categories, reflecting their origin from discrete B-cell differentiation stages[1] or the co-regulated expression of comprehensive transcriptional signatures[4]. The cell-of-origin classification schema comprises GCB-DLBCL, derived from a germinal centre (GC) centroblast; the less curable ABC-DLBCL, the expression pattern of which resembles that of cells committed to plasmacytic differentiation; primary mediastinal large B-cell lymphomas (PMBL), arising from thymic B cells[5]; and cases that remain unclassified (NC)[1,6]. A key feature of ABC-DLBCL is the activation of the NF-κB signalling pathway, as shown by the preferential expression of known NF-κB target genes and the dependence of ABC-DLBCL cell lines on NF-κB activity for proliferation and survival[2,7]. A recent study reported that ~8% of ABC-DLBCL carry oncogenic mutations of CARD11 (ref. 3), a cytoplasmic scaffolding protein required for activation of NF-κB during antigen-dependent signalling[8]. However, the molecular mechanism underlying NF-κB

activation in the remaining large fraction of cases remains unknown, leaving open the possibility that it may reflect a physiological status of the normal ABC-DLBCL counterpart.

To address this issue, we first characterized 168 DLBCL samples, representative of major subtypes, for the presence of active NF-κB complexes by using immunohistochemical assays detecting nuclear NFKB1 (also called p50; read-out for the classical pathway) and NFKB2 (p52; alternative pathway)[9,10] (Fig. 1a). Nuclear localization of NF-κB was observed in the tumour cells of 61% of ABC-DLBCL and a smaller fraction (30%) of GCB-DLBCL, as well as in 3 out of 9 unclassified and 36 out of 73 not profiled DLBCL (Fig. 1b). Both classical and alternative NF-κB pathways were found to be involved, occasionally within the same sample (one-third of the positive cases), and consistent with the established role of specific signals (for example, CD40–CD40L) in the activation of both pathways[11,12]. Engagement of the alternative NF-κB pathway was also documented by detection of p52, the active product of p100 processing, in western blot assays (Fig. 1c). Gene set enrichment analysis (GSEA) of transcriptionally profiled cases confirmed that the gene expression signature of ABC-DLBCL is significantly enriched in NF-κB target genes (Supplementary Table 1) with respect to both normal GC centroblasts, used as negative control[13] ($P < 0.005$; not shown), and GCB-DLBCL ($P = 0.03$; Fig. 1d). Moreover, all immunohistochemistry-positive samples displayed a transcriptional signature of NF-κB pathway activity. The fraction of cases presenting high NF-κB transcriptional activity by GSEA was higher than that defined by immunohistochemistry (>95% ABC-DLBCL and ~47% GCB-DLBCL; Fig. 1e, f). This difference probably reflects the higher sensitivity of gene-expression-profile-based approaches, but also their inability to discriminate signals deriving from infiltrating reactive cells. Thus, immunohistochemistry may provide a rapid and specific, although relatively less sensitive, approach for the identification of constitutively active NF-κB on routine diagnostic material. Both methods revealed that NF-κB signalling is not limited to ABC-DLBCL, but may also be present in a smaller subset of GCB-DLBCL.

To investigate whether constitutive NF-κB activation in ABC-DLBCL represents a primary pathogenetic event or reflects the intrinsic program of the tumour cell of origin, we screened for mutations the complete coding sequence of 31 NF-κB-pathway genes in 14 samples (Supplementary Table 2). Genes found mutated after filtering for known polymorphisms and synonymous mutations were further analysed in a validation panel composed of 87 DLBCL (23 ABC, 44 GCB and 20 unclassified/non-GC; Supplementary Fig. 1).

This strategy identified a total of 48 sequence changes distributed in 6 different genes, including the NF-κB negative regulator *A20*

---

[1]Institute for Cancer Genetics and the Herbert Irving Comprehensive Cancer Center, [2]Joint Centers for Systems Biology, [3]Department of Pathology & Cell Biology, Columbia University, New York, New York 10032, USA. [4]Laboratory of Experimental Oncology and Lymphoma Unit, Oncology Institute of Southern Switzerland (IOSI), 6500 Bellinzona, Switzerland. [5]Pathology Unit, Unit of Lymphoid Malignancies, San Raffaele Scientific Institute, 20132 Milan, Italy. [6]Department of Pathology and Laboratory Medicine, Weill Medical College of Cornell University, New York, New York 10021, USA. [7]Department of Genetics & Development, Columbia University, New York, New York 10032, USA.
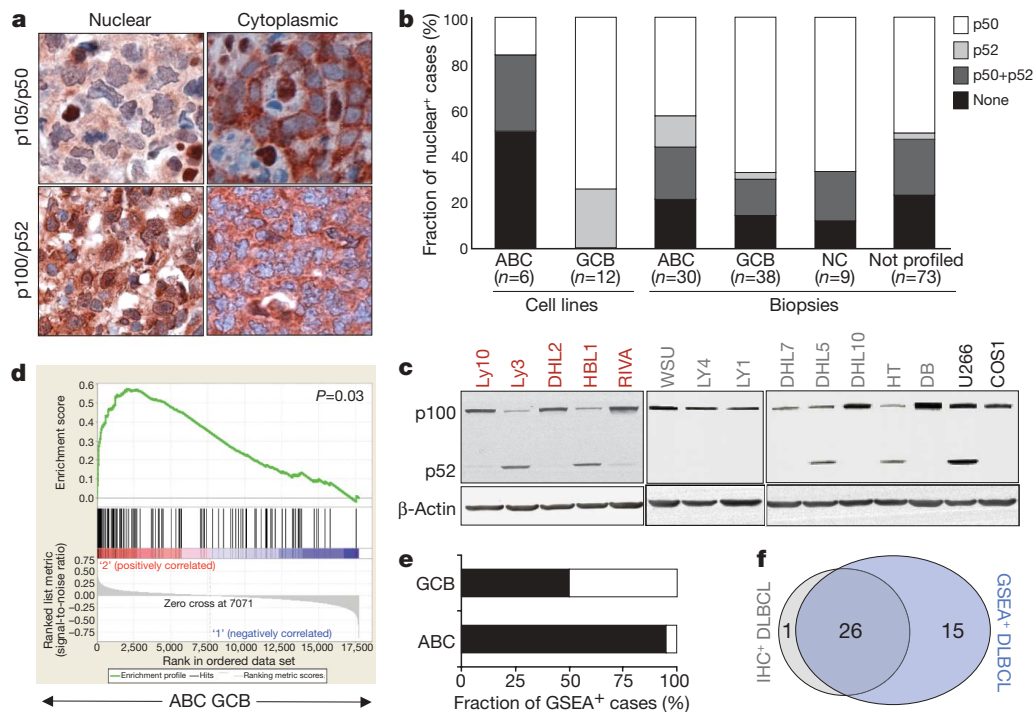
**Figure 1 | The NF-κB pathway is active in ABC-DLBCL and in a smaller fraction of GCB-DLBCL. a,** Immunohistochemical staining of DLBCL biopsies with anti-NFKB1 (p105/p50, top) and anti-NFKB2 (p100/p52, bottom) antibodies. Nuclear localization of NF-κB denotes active signalling, as opposed to the inactive, cytoplasmic complex. Original magnification, ×200. **b,** Prevalence of cases displaying constitutive NF-κB activation in DLBCL subgroups. Colour codes indicate nuclear p50, p52 or both. **c,** Western blot analysis of DLBCL cell lines for processing of p100 to p52 (red, ABC-DLBCL; grey, GCB-DLBCL). The multiple myeloma cell line U266 and the epithelial cell line COS are used as positive and negative controls, respectively. Actin, loading control. **d,** GSEA enrichment score and distribution of NF-κB target genes along the rank of transcripts differentially expressed in ABC-DLBCL versus GCB-DLBCL. **e,** Percentage of samples showing a transcriptional NF-κB signature by GSEA. **f,** Venn diagram illustrating the overlap between immunohistochemistry-defined and GSEA-defined NF-κB-positive cases.

(*TNFAIP3*)[14–16] and the positive regulators *CARD11* (ref. 8), *TNFRSF11A* (*RANK*)[17], *TRAF2* (ref. 18), *TRAF5* (ref. 19) and *MAP3K7* (*TAK1*)[20] (Table 1 and Supplementary Table 3). Mutations were preferentially associated with the ABC-DLBCL phenotype, where 51.3% of the samples analysed showed alterations in one or more gene, compared with 22.7% GCB-DLBCL (Table 1 and Supplementary Table 4). In addition, 7 out of 20 (35%) non-GC DLBCLs were found to be mutated. Analysis of paired normal DNA, available from 8 samples, indicated the somatic origin of these events in at least one sample/gene.

The most commonly affected gene was *A20*, which codes for a dual function ubiquitin-modifying enzyme belonging to the ovarian tumour (OTU) domain-containing family of deubiquitinating enzymes and required for termination of NF-κB responses in the classical NF-κB pathway[14–16]. Notably, the *A20* locus is positioned on chromosomal band 6q23.3, a region frequently deleted in aggressive B-cell lymphomas, and suggested to contain a tumour suppressor[21,22].

We therefore examined this gene in 68 additional DLBCL biopsies, immunohistohemically classified as GC and non-GC based on the Hans algorithm, with minor modifications (see Methods)[23]. Combined, the two screenings led to the identification of 26 mutational events, distributed in 22 cases and almost exclusively segregating with an ABC/non-GC phenotype (9 out of 37 ABC-DLBCL and 10 out of 51 non-GC/NC-DLBCL, versus 2 out of 72 GCB-DLBCL; Fig. 2a). Sequence changes included nonsense mutations introducing premature termination codons ($n = 12$); frameshift deletions/insertions ($n = 7$ and 5, respectively); and nucleotide substitutions at consensus splice donor sites ($n = 2$), which were documented by cDNA amplification and sequencing to generate aberrant transcripts that retain intronic sequences and have lost their coding potential (Fig. 2b and Supplementary Table 5). The common consequence of these mutations is the production of severely truncated A20 polypeptides that lack functionally relevant domains (Supplementary Fig. 2) and are either unstable or functionally impaired, as experimentally demonstrated in transient transfection/NF-κB reporter gene assays (Supplementary Fig. 3)[14,16].

In four samples, each displaying two mutational events, sequencing analysis of *A20* transcripts after cDNA amplification and cloning demonstrated that the mutations were located on separate alleles, leading to biallelic gene inactivation. Moreover, fluorescence *in situ* hybridization (FISH) analysis using specific probes and/or direct sequencing revealed deletion of the second allele in 12 out of 14 mutated cases with available material (Fig. 2c, d and Supplementary Table 5). Homozygous *A20* deletions were found in seven additional cases, one of which harboured a focal deletion (<420 kilobases (kb)) encompassing *A20* and *OLIG3*, a gene not expressed in B cells (Supplementary Fig. 4 and Supplementary Table 6), providing strong evidence for *A20* being the target of the lesion. In all samples, loss of the signal accounted for ≥90% of the tumour cell population,

**Table 1 | NF-κB pathway genes found mutated in DLBCL**

| Gene symbol (synonym) | Number of mutated/tested cases (%) | | |
|---|---|---|---|
| | ABC-DLBCL | GCB-DLBCL | Non-GC/NC-DLBCL |
| *TNFAIP3* (*A20*) | 9/37 (24.3) | 1/44 (2.3) | 4/20 (20) |
| *CARD11* | 4/37 (10.8) | 3/44 (6.8)† | 2/20 (10) |
| *TNFRSF11A* (*RANK*) | 3/37 (8.1)* | 1/44 (2.3) | 2/20 (10) |
| *TRAF5* | 2/37 (5.4) | 2/44 (4.5) | 1/20 (5) |
| *TRAF2* | 1/37 (2.7) | 4/44 (9.1)‡ | 1/20 (5) |
| *MAP3K7* (*TAK1*) | 2/37 (5.4) | 0/44 | 0/20 |
| All genes | 19/37 (51.3) | 10/44 (22.7)§ | 7/20 (35) |

* One additional cell line was found to carry a mutation in a minority of the population.
† Two of the three mutated samples are cell lines; analysis restricted to exons 4–9, encoding the coiled-coil domain.
‡ Three of the four mutated samples are cell lines.
§ Four of the ten mutated samples are cell lines.

718
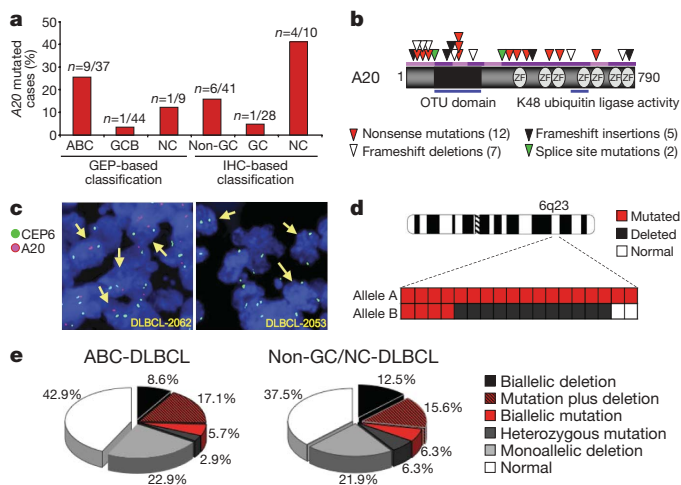
**Figure 2 | Mutations and deletions of the *A20* gene in ABC-DLBCL.**
**a**, Percentage of *A20* mutated cases in various DLBCL subtypes. The exact number over total analysed is given at the top of each column. NC, samples co-expressing CD10 and IRF4 (see Methods). GEP, gene expression profile.
**b**, Schematic representation of the human A20 protein, with its key functional domains. Colour-coded symbols depict distinct alterations leading to A20 inactivation. OTU, ovarian tumour domain; ZF, A20 zinc-finger domains. **c**, Dual-colour FISH analysis of representative DLBCLs, hybridized with A20-specific probes (red) and a chromosome 6 centromeric probe (green). Arrows point to representative cells displaying hemizygous (left panel) or homozygous (right panel) *A20* deletions. **d**, Chromosome 6 ideogram; the region encompassing *A20* on band 6q23 is enlarged in the bottom panel, where columns represent individual cases, and rows correspond to the two alleles. **e**, Overall frequency of *A20* structural alterations (inactivating mutations and deletions, combined).

consistent with a clonally represented event (Supplementary Table 7). Thus, 32% of ABC-DLBCL and ~34% of non-GC/NC-DLBCL have lost both copies of the *A20* gene due to the presence of inactivating

mutations and/or deletions (Fig. 2e). Notably, monoallelic deletions were also observed in 23% of ABC-DLBCL and ~22% of non-GC-DLBCL. Because expression of the wild-type allele was still detected in the three cell lines investigated, these data may suggest haplo-insufficiency or the involvement of a second gene in the context of larger 6q chromosomal deletions, frequently observed in aggressive lymphomas[21]. Collectively, these findings indicate that *A20* is frequently inactivated in DLBCL by a two-hit mechanism typical of tumour suppressor genes.

To test directly the role of A20 in cell transformation, we used lentiviral expression vectors and reintroduced A20 in two cell lines (SUDHL2 and RC-K8) carrying biallelic *A20* gene inactivation. As shown in Fig. 3a–c, A20 reconstitution induced apoptosis and cell growth arrest in the A20-null cell lines, but not in two control lines carrying an intact *A20* locus and lacking constitutive NF-κB activity. Consistently, fluorescence-activated cell sorting (FACS) analysis of green fluorescent protein (GFP) expression documented the progressive disappearance of the A20-positive population (identified by GFP) in SUDHL2 and RC-K8, as opposed to SUDHL4 and SUDHL7 or to empty-vector-transduced cells, where >90% of the population was GFP[+] 8 days after sorting (Fig. 3d). Notably, most A20-reconstituted cells showed complete cytoplasmic relocation of p50 by immunofluorescence staining (Fig. 3e), indicating an A20-dependent block in NF-κB signalling and consistent with its well established role in the termination of NF-κB responses *in vitro* and *in vivo*[14–16]. Together, these findings strongly suggest a tumour suppressor role for A20, the loss of which may contribute to DLBCL pathogenesis by causing supra-physiological activation of NF-κB which, in turn, has oncogenic properties via inhibiting apoptosis and promoting cell proliferation[24].

Less commonly, missense mutations were found in positive regulators of the NF-κB pathway, namely the scaffolding proteins CARD11 (11%), TRAF2 (3%) and TRAF5 (5%), which mediate NF-κB activation via oligomerization and activation of the IKK kinase; the MAP3K7 serine-threonine kinase (5%), which directly
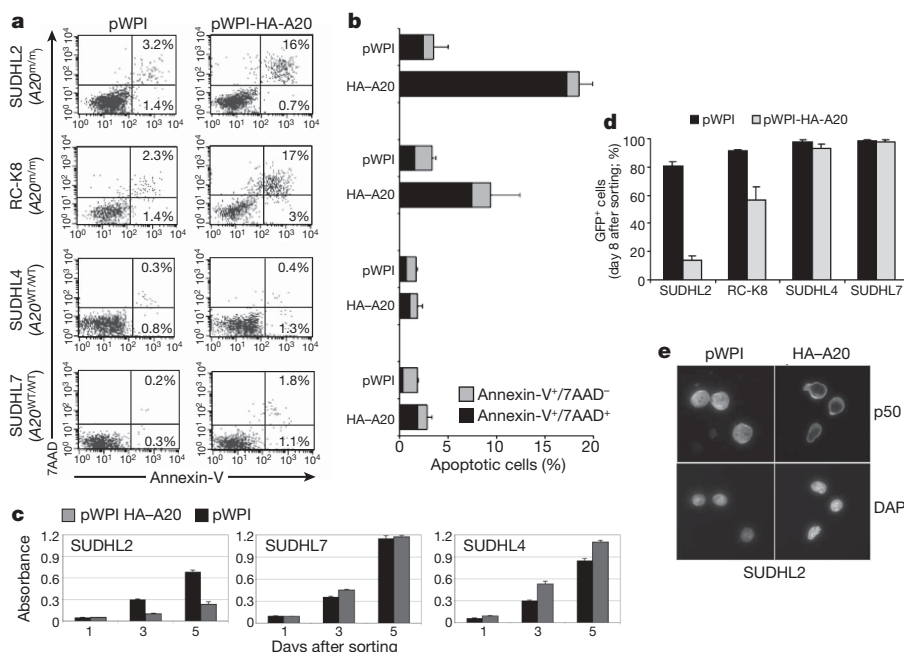


**Figure 3 | Reconstitution of *A20*-null cell lines causes apoptosis and cell growth arrest via inhibition of NF-κB.** **a**, Flow cytometric analysis of Annexin-V-phycoerythrin (PE)/7-amino-actimomycin D (7AAD) staining in *A20*-null and wild-type cell lines, transduced with the indicated vectors; numbers indicate the percentage of single- and double-positive cells (lower and upper right quadrants). The results of three independent experiments are summarized in **b** (mean and s.d.). **c**, Reduced viability in SUDHL2 cells complemented with A20 expression vectors (grey bars), as compared to

vector alone (black bars; mean and s.d., *n* = 3). No effect was observed in two control cell lines. **d**, Analysis of GFP expression after transduction with pWPI and pWPI-HA-A20 documents the disappearance of the GFP[+] population in SUDHL2[HA–A20] and RC-K8[HA–A20] cell lines, but not in SUDHL2[pWPI] and RC-K8[pWPI] or in A20 wild-type cell lines (mean and s.d., *n* = 2). **e**, Immunofluorescence analysis of p105/p50 in SUDHL2 cells transduced with pWPI and pWPI-HA-A20 (top). Nuclei are identified by 4′,6′-diamidino-2-phenylindole (DAPI, bottom).

phosphorylates IKK[20,25]; and the cell-surface receptor TNFRSF11A (8.1%), involved in classical NF-κB responses (Table 1 and Supplementary Table 8). Notably, single nucleotide polymorphism (SNP)-array data showed amplification of the regions harbouring these genes in 41 cases, suggesting their possible dominant role in activating NF-κB (not shown). To investigate the functional significance of these mutations, we examined their ability to activate a luciferase reporter vector driven by two NF-κB-responsive elements in transient transfection assays. In agreement with a recent study[3], CARD11 mutations potentiate its NF-κB transactivation activity, in the absence of further stimuli (Supplementary Fig. 6a, b). Significantly enhanced NF-κB activity was also observed on transfection of the ABC-DLBCL-derived TRAF2(P186R) mutant (Supplementary Fig. 6c). When expressed in the DLBCL cell line SUDHL6, which lacks constitutive NF-κB activity, this mutant was sufficient to induce nuclear p50 translocation in most cells, indicating its ability to stimulate this pathway *in vivo* (Supplementary Fig. 6d). Conversely, no significant differences were associated with four GCB-derived *TRAF2* mutant alleles and with the mutant *MAP3K7* allele (not shown). Because these mutations were mostly observed in cell lines, their somatic origin could not be verified, leaving open the possibility that they represent polymorphisms that have not been reported previously. Alternatively, these data may suggest a more subtle effect of the mutations, not detectable by the experimental approach used. Although further studies will be required to dissect the significance of these alterations *in vivo*, our data show that at least 15 out of 37 (40.5%) ABC-DLBCL (those with *A20*, *CARD11* and *TRAF2* alterations) display mutations of proven functional significance in activating NF-κB.

The identification of multiple genetic alterations converging on the same pathway in a sizable fraction of ABC-DLBCL provides a genetic explanation for the presence of constitutive NF-κB activity in this tumour type, indicating a role for this signalling pathway as a primary pathogenetic event in lymphomagenesis. The most prominent player in this scenario is the known NF-κB negative regulator A20. Notably, structural alterations affecting this gene are also found in Hodgkin's lymphoma, PMBL and marginal zone lymphoma[26,27,28]. These findings, together with the evidence of its functional role in modulating NF-κB[14–16], identify *A20* as a relevant tumour suppressor gene, the inactivation of which may contribute to the pathogenesis of several lymphoma subtypes. Because A20 is itself a target of NF-κB and needs to be induced in order to exert its negative feedback effect, additional upstream events are probably required by the tumour cells to activate this signalling cascade and promote selective pressure for A20 inactivation, including engagement of the B-cell receptor by the antigen, CD40–CD40L signalling and BAFF–BAFFR interaction. However, the observation that, in some cases, multiple genes are simultaneously altered within this signalling pathway, as a combination of positive and negative regulators (for example, *A20* and *TNFRSF11A* or *MAP3K7*), suggests that additional upstream genetic lesions may complement the loss of *A20*.

Constitutive NF-κB activation may promote malignant transformation by providing anti-apoptotic and pro-proliferative signals. Notably, these lesions occur in the same cases displaying structural alterations of BCL6 and BLIMP1[29,30], which may contribute to lymphomagenesis by suppressing genotoxic responses (BCL6)[31] and/or preventing terminal B-cell differentiation (BCL6, BLIMP1)[30]. As such, these findings provide the rationale and the assays for the identification of DLBCL patients potentially benefiting from targeted anti-NF-κB therapeutic approaches.

## METHODS SUMMARY

**Characterization of NF-κB activity.** The presence of active NF-κB complexes in DLBCL cell lines and primary biopsies was analysed by immunohistochemistry and immunofluorescence analysis of paraffin-embedded tissue sections and cytospin preparations using anti-p105/p50 and anti-p100/p52 antibodies, and by GSEA of NF-κB target genes on Affymetrix U133Plus_2 gene expression profile data.

**Mutation analysis.** The complete coding sequences and exon/intron junctions of 31 NF-κB genes were analysed by PCR amplification and direct sequencing of genomic DNA as described[30]. Mutations were confirmed by sequencing of both strands on independent PCR products, whereas previously reported polymorphisms, changes present in matched normal DNA and silent mutations were filtered from the analysis.

**FISH analysis.** FISH analysis was performed on tissue microarrays using two specific BAC probes spanning the *A20* gene and a centromeric probe for chromosome 6 (ref. 30).

**Lentiviral transduction of A20 expression constructs.** For the reconstitution assay, DLBCL cell lines were transduced with lentiviral vectors expressing GFP alone (pWPI) or wild-type human A20 linked to IRES-GFP (pWPI-HA-A20), and analysed for effects on survival, cell growth and NF-κB activity. Productively transduced (GFP+) cells were purified by cell sorting before use in proliferation assays and immunofluorescence staining of nuclear p50.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

1.  Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
2.  Davis, R. E., Brown, K. D., Siebenlist, U. & Staudt, L. M. Constitutive nuclear factor κB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. *J. Exp. Med.* **194**, 1861–1874 (2001).
3.  Lenz, G. *et al.* Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. *Science* **319**, 1676–1679 (2008).
4.  Monti, S. *et al.* Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood* **105**, 1851–1861 (2005).
5.  Savage, K. J. *et al.* The molecular signature of mediastinal large B-cell lymphoma differs from that of other diffuse large B-cell lymphomas and shares features with classical Hodgkin lymphoma. *Blood* **102**, 3871–3879 (2003).
6.  Wright, G. *et al.* A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc. Natl Acad. Sci. USA* **100**, 9991–9996 (2003).
7.  Ngo, V. N. *et al.* A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* **441**, 106–110 (2006).
8.  Rawlings, D. J., Sommer, K. & Moreno-Garcia, M. E. The CARMA1 signalosome links the signalling machinery of adaptive and innate immunity in lymphocytes. *Nature Rev. Immunol.* **6**, 799–812 (2006).
9.  Li, Q. & Verma, I. M. NF-κB regulation in the immune system. *Nature Rev. Immunol.* **2**, 725–734 (2002).
10. Hoffmann, A. & Baltimore, D. Circuitry of nuclear factor κB signaling. *Immunol. Rev.* **210**, 171–186 (2006).
11. Pomerantz, J. L. & Baltimore, D. Two pathways to NF-κB. *Mol. Cell* **10**, 693–695 (2002).
12. Coope, H. J. *et al.* CD40 regulates the processing of NF-κB2 p100 to p52. *EMBO J.* **21**, 5375–5385 (2002).
13. Basso, K. *et al.* Tracking CD40 signaling during germinal center development. *Blood* **104**, 4088–4096 (2004).
14. Boone, D. L. *et al.* The ubiquitin-modifying enzyme A20 is required for termination of Toll-like receptor responses. *Nature Immunol.* **5**, 1052–1060 (2004).
15. Lee, E. G. *et al.* Failure to regulate TNF-induced NF-κB and cell death responses in A20-deficient mice. *Science* **289**, 2350–2354 (2000).
16. Wertz, I. E. *et al.* De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF-κB signalling. *Nature* **430**, 694–699 (2004).
17. Anderson, D. M. *et al.* A homologue of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. *Nature* **390**, 175–179 (1997).
18. Rothe, M., Sarma, V., Dixit, V. M. & Goeddel, D. V. TRAF2-mediated activation of NF-κB by TNF receptor 2 and CD40. *Science* **269**, 1424–1427 (1995).
19. Rothe, M., Wong, S. C., Henzel, W. J. & Goeddel, D. V. A novel family of putative signal transducers associated with the cytoplasmic domain of the 75 kDa tumor necrosis factor receptor. *Cell* **78**, 681–692 (1994).
20. Yamaguchi, K. *et al.* Identification of a member of the MAPKKK family as a potential mediator of TGF-β signal transduction. *Science* **270**, 2008–2011 (1995).
21. Gaidano, G. *et al.* Deletions involving two distinct regions of 6q in B-cell non-Hodgkin lymphoma. *Blood* **80**, 1781–1787 (1992).
22. Honma, K. *et al.* TNFAIP3 is the target gene of chromosome band 6q23.3-q24.1 loss in ocular adnexal marginal zone B cell lymphoma. *Genes Chromosom. Cancer* **47**, 1–7 (2008).
23. Hans, C. P. *et al.* Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood* **103**, 275–282 (2004).
24. Karin, M. Nuclear factor-κB in cancer development and progression. *Nature* **441**, 431–436 (2006).
25. Wang, C. *et al.* TAK1 is a ubiquitin-dependent kinase of MKK and IKK. *Nature* **412**, 346–351 (2001).

26. Novak, U. *et al.* The NF-κB negative regulator *TNFAIP3* (A20) is inactivated by somatic mutations and genomic deletions in marginal zone B-cell lymphomas *Blood* doi:10.1182/blood-2008-08-174110 (2009).

27. Kato, M. *et al.* Frequent inactivation of A20 in B-cell lymphomas. *Nature* doi:10.1038/nature07969 (this issue).

28. Schmitz, R. *et al.* TNFAIP3 (A20) is a tumor suppressor gene in Hodgkin lymphoma and primary mediastinal B cell lymphoma. *J. Exp. Med.* (in the press).

29. Saito, M. *et al.* A signaling pathway mediating downregulation of BCL6 in germinal center B cells is blocked by BCL6 gene alterations in B cell lymphoma. *Cancer Cell* **12**, 280–292 (2007).

30. Pasqualucci, L. *et al.* Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. *J. Exp. Med.* **203**, 311–317 (2006).

31. Phan, R. T. & Dalla-Favera, R. The *BCL6* proto-oncogene suppresses p53 expression in germinal-centre B cells. *Nature* **432**, 635–639 (2004).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Author Contributions** L.P. and R.D.-F. designed the study. M.C., A.G. and Q.S. performed experiments; L.P. performed the A20 functional assays; W.K.L. and A.Ca. developed tools for genome-wide expression profile analysis; S.V.N. performed the FISH analysis; A.Ch. and G.B. analysed all immunohistochemistry data; A.Ch., G.B., F.B. and M.P. provided DLBCL samples; M.B., F.B. and M.S. performed SNP array analysis; L.P. designed experiments, coordinated the study, analysed data and wrote the manuscript, which was commented on by all authors.

**Author Information** The expression data and the 250K SNP data reported here have been deposited in the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo) database (series accession number GSE12195 and GSE15127). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.P. (lp171@columbia.edu).

## METHODS

**DNA extraction, amplification and sequencing.** Genomic DNA was extracted according to standard methods. In eight cases with available matched non-neoplastic tissue, DNA was also extracted from paraffin-embedded material using the QIAamp DNA mini Kit (QIAGEN). Sequences for all annotated exons and flanking introns of the 31 NF-κB pathway genes listed in Supplementary Table 2 were obtained from the UCSC Human Genome database, using the corresponding mRNA accession number as a reference. The Primer 3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) was used to design oligonucleotides for amplification and sequencing of each coding exon (plus ~50 bp of adjacent introns), available on request. The primers used for analysis of the six genes found mutated are reported in Supplementary Table 9. Purified amplicons were sequenced directly from both strands as described, and compared to the corresponding germline sequences, using the Mutation Surveyor Version 2.41 software package (Soft Genetics LLC)[30]. Synonymous mutations, changes due to previously reported polymorphisms (Human dbSNP Database at NCBI, Build 129, and Ensembl database) and changes present in normal DNA from the same patient, when available, were excluded. Somatic mutations were confirmed on independent PCR products. In cases displaying more than one event within a single gene (*A20*, *CARD11* and *TRAF5*), the allelic distribution of the mutations was determined by cloning and sequencing full-length PCR products obtained from cDNA ($n = 10$ clones each)[32].

**Tissue microarrays, immunohistochemistry and immunofluorescence staining.** The construction of the DLBCL tissue microarray was performed according to standard procedures and the protocols for immunohistochemical and immunofluorescence staining are described in ref. 13. Samples were classified as GC or non-GC type based on expression of CD10, BCL6 and IRF4 according to ref. 23, except that the rare CD10[+] cases co-expressing IRF4 were designated as non-classified (NC), as IRF4 is a known marker of B-cell activation, and is normally absent in BCL6[+] GC centroblasts[33]. The percentage and staining intensity of neoplastic B-cells were independently scored by two pathologists (A.Ch. and G.B.), using a cut-off of 30% positive cells. Cases were considered to be positive for NF-κB activity when ≥30% of tumour cells showed nuclear NF-κB localization. The antibodies used were rabbit monoclonal anti NF-κB1 p105/p50 and NF-κB2 p100/p52 (18D10) (Cell Signaling Technology).

**RNA extraction, generation of gene expression profiles and DLBCL classification.** The protocols for RNA extraction, cRNA labelling and hybridization to Affymetrix GeneChip U133Plus_2 microarrays are described in detail in ref. 13. Gene expression data were normalized by the MAS 5.0 software, followed by log$_2$ transformation. The DLBCL primary biopsies were classified into GCB ($n = 38$), ABC ($n = 30$) and unclassified ($n = 9$) as previously described[6], using a linear predictive score and 22 of the 27 original lymphochip predictor genes which were represented in the U133Plus_2 array and showed the best *t*-statistics. Cases displaying inconsistencies between COO-classification, unsupervised hierarchical clustering analysis and immunohistochemistry-based classification were considered as unclassified or excluded from further gene expression profile based analyses.

**Gene set enrichment analysis (GSEA).** Enrichment analysis of NF-κB target genes was performed as previously described[34] using the genes listed in Supplementary Table 1 and gene expression profiles from the DLBCL biopsies (GSEA v2.0 at http://www.broad.mit.edu/gsea). The NF-κB target gene set was generated by combining previously reported target genes identified in genome-wide expression profile studies of B cells, and included genes that were specifically downregulated after genetic (induction of NF-κB super-repressor; *CARD11* shRNA) or pharmacological (IKK inhibitors) manipulation of NF-κB in representative ABC-DLBCL and Hodgkin's lymphoma cell lines. GSEA was also used to assess whether individual DLBCL samples expressed a transcriptional signature of NF-κB activation. To this end, the expression of each gene on the U133Plus_2 microarray was first converted into a *z*-score using ten samples

of purified normal GC B cells as a baseline. Genes were then ranked by their *z*-score from the most positive to the most negative value, and the 120 genes of the NF-κB gene set were intersected with the ordered list to compute GSEA enrichment scores. The algorithm was set to implement weighted scoring scheme and the enrichment score significance was assessed by 100,000 permutation tests. Samples attaining significant *P*-value ($P < 0.05$, Bonferroni corrected) were designated as samples with activated NF-κB.

**FISH.** Two PAC clones (RP11-703G8 and RP1-702P5) spanning the *A20* gene were obtained from BACPAC Resources at http://bacpac.chori.org. DNA was labelled by nick translation using spectrum orange dUTP fluorochrome (Vysis). A Spectrum green-labelled centromeric probe (Vysis) was used to enumerate chromosome 6. Paraffin-embedded tissue sections from TMAs were baked overnight at 60 °C and processed using a paraffin pre-treatment kit (Vysis). FISH was performed on DAPI-stained slides by standard methods, and hybridization signals were scored on at least 500 interphase nuclei/core (that is, five representative areas with at least 100 nuclei each). Slides were evaluated for probe signal intensity and signal to background ratio. As control, multiple sections from normal tonsils were included in each TMA. Normal variation corresponded to $9.7 \pm 4.6\%$ of the nucleated cells for loss of 6q23 signal, and $33.5 \pm 12.5$ for monosomy 6. Cases were diagnosed as positive when the fraction of cells showing an abnormal pattern was above the mean $+2$ s.d. ($+1$ s.d. for monosomies). The percentage of tumour cells in each core was estimated by histological analysis of serial TMA sections.

**Lentiviral infections.** The replication-deficient lentiviral expression construct pWPI-HA-A20 was generated by subcloning the full-length *A20* cDNA sequences into the PmeI restriction site of pWPI, in front of IRES-GFP. Viral supernatants were obtained by co-transfecting 293T cells with the lentiviral expression vectors and vectors expressing the helper virus Δ8.9 and the VSV-G envelope glycoprotein[35,36]. Conditioned medium was harvested over 48–62 h and used directly to infect the indicated cell lines according to standard methods. Transduction efficiencies were determined by FACS analysis of GFP expression after 48–72 h. For cell proliferation assays, western blot analysis of exogenous A20 expression, and immunofluorescence staining of nuclear p105/p50, productively infected cells were sorted by flow cytometry on a BD FACSAria Cell Sorter, based on GFP expression.

**Apoptosis and proliferation assays.** The effect of A20 expression on cell survival was measured 48 and 72 h after infection by flow cytometric analysis of Annexin-V-PE and 7-amino-actimomycin D (7AAD; BD Pharmingen Biosciences) stained cells, gating on the GFP[+] population. Data were acquired on a FACSCalibur (Becton Dickinson) and analysed with the CELLQuest software. Cell proliferation was monitored on sorted GFP[+] cells using the MTT reagent (Roche) according to the manufacturer's instructions. The fraction of live GFP[+] cells was also measured over time by FACS analysis, and compared to the initial GFP[+] population (determined 2 days after lentiviral transduction) on at least three independent experiments.

32. Pasqualucci, L. *et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412**, 341–346 (2001).
33. Falini, B. *et al.* A monoclonal antibody (MUM1p) detects expression of the MUM1/IRF4 protein in a subset of germinal center B cells, plasma cells, and activated T cells. *Blood* **95**, 2084–2092 (2000).
34. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
35. Lois, C., Hong, E. J., Pease, S., Brown, E. J. & Baltimore, D. Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science* **295**, 868–872 (2002).
36. Naldini, L. *et al. In vivo* gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272**, 263–267 (1996).

# LETTERS

# F-box protein FBXO31 mediates cyclin D1 degradation to induce G1 arrest after DNA damage

Manas K. Santra[1]*, Narendra Wajapeyee[1]* & Michael R. Green[1]

In response to DNA damage, eukaryotic cells initiate a complex signalling pathway, termed the DNA damage response (DDR), which coordinates cell cycle arrest with DNA repair. Studies have shown that oncogene-induced senescence, which provides a barrier to tumour development, involves activation of the DDR[1–3]. Using a genome-wide RNA interference (RNAi) screen, we have identified 17 factors required for oncogenic BRAF to induce senescence in primary fibroblasts and melanocytes[4]. One of these factors is an F-box protein, FBXO31, a candidate tumour suppressor encoded in 16q24.3, a region in which there is loss of heterozygosity in breast, ovarian, hepatocellular and prostate cancers[5–9]. Here we study the cellular role of FBXO31, identify its target substrate and determine the basis for its growth inhibitory activity. We show that ectopic expression of FBXO31 acts through a proteasome-directed pathway to mediate the degradation of cyclin D1, an important regulator of progression from G1 to S phase, resulting in arrest in G1. Cyclin D1 degradation results from a direct interaction with FBXO31 and is dependent on the F-box motif of FBXO31 and phosphorylation of cyclin D1 at Thr 286, which is known to be required for cyclin D1 proteolysis. The involvement of the DDR in oncogene-induced senescence prompted us to investigate the role of FBXO31 in DNA repair. We find that DNA damage induced by γ-irradiation results in increased FBXO31 levels, which requires phosphorylation of FBXO31 by the DDR-initiating kinase ATM. RNAi-mediated knockdown of FBXO31 prevents cells from undergoing efficient arrest in G1 after γ-irradiation and markedly increases sensitivity to DNA damage. Finally, we show that a variety of DNA damaging agents all result in a large increase in FBXO31 levels, indicating that induction of FBXO31 is a general response to genotoxic stress. Our results reveal FBXO31 as a regulator of the G1/S transition that is specifically required for DNA damage-induced growth arrest.

We studied FBXO31 function in human SK-MEL-28 melanoma cells. We first tested whether, as in breast cancer cell lines[9], ectopic expression of FBXO31 could induce G1 arrest in SK-MEL-28 cells. Fluorescence-activated cell sorting (FACS) analysis revealed, as expected, that treatment of SK-MEL-28 cells with nocodazole resulted in G2/M arrest (Fig. 1a). Ectopic expression of FBXO31 prevented the nocodazole-induced G2/M block and instead substantially increased the fraction of cells in G1. Identical results were obtained in cells ectopically expressing FBXO31 in the absence of nocodazole (Supplementary Fig. 1). Moreover, ectopic expression of FBXO31 blocked DNA synthesis (Fig. 1b). Notably, ectopic expression of FBXO31 also markedly inhibited the growth of SK-MEL-28 cells in culture (Supplementary Fig. 2a) and SK-MEL-28 mouse xenografts (Supplementary Fig. 2b).

F-box proteins are best known for their role as the substrate-recognition components of the SCF (SKP/Cullin/F-box protein) class of E3

ubiquitin ligases[10]. To investigate the basis by which FBXO31 induces G1 arrest, we analysed a panel of cell cycle regulatory proteins by immunoblot analysis after ectopic expression of FBXO31. Figure 1c shows that after expression of FBXO31, but not vector alone (Supplementary Fig. 3), the level of cyclin D1, a key regulator of the G1/S-phase transition[11], markedly declined. In contrast, the levels of other G1 cyclins and the G2/M regulator cyclin B1 were unaffected.
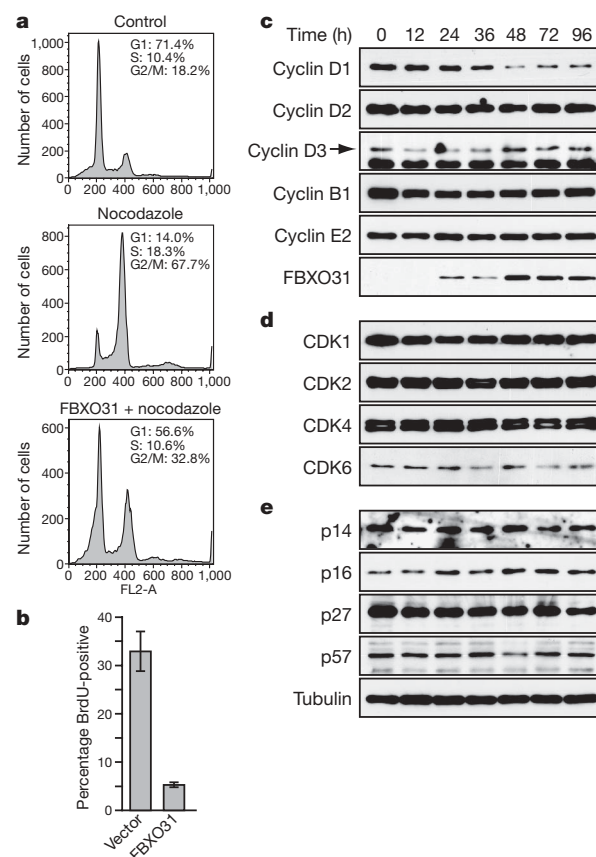


**Figure 1 | Ectopic expression of FBXO31 induces G1 arrest and selective degradation of cyclin D1. a**, FACS analysis in SK-MEL-28 cells transduced with a retrovirus expressing empty vector or FBXO31 in the absence or presence of nocodazole. **b**, DNA replication assay, monitored by BrdU incorporation (error bars indicate s.d.; $n = 3$). **c–e**, Immunoblot analysis showing levels of cyclins (**c**), CDKs (**d**) and CDK inhibitors (**e**) at various time points after transduction with an FBXO31 retrovirus. Tubulin was monitored as a loading control. Ectopic expression of FBXO31 did not induce a DDR (Supplementary Fig. 17).

[1]Howard Hughes Medical Institute, Programs in Gene Function and Expression and Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.
*These authors contributed equally to this work.

Ectopic FBXO31 expression also had no effect on the levels of all cyclin-dependent kinases (CDKs) (Fig. 1d) and CDK inhibitors (Fig. 1e) examined. As expected, the decrease in cyclin D1 levels was accompanied by an increase in the fraction of cells in G1 (Supplementary Fig. 4). RNAi-mediated knockdown of cyclin D1 in SK-MEL-28 cells resulted in a similar G1 arrest (Supplementary Fig. 5). Collectively, these results suggest that FBXO31 induces G1 arrest through the selective degradation of cyclin D1.

We next sought to determine whether the FBXO31-mediated decrease in cyclin D1 resulted from proteasomal degradation. Figure 2a shows that addition of the proteasome inhibitor lactacystin blocked the ability of ectopically expressed FBXO31 to decrease cyclin D1 levels. Consistent with this result was our observation that levels of *CCND1* messenger RNA (encoding cyclin D1) were unaffected by ectopic FBXO31 expression (Fig. 2b). Moreover, after ectopic expression of FBXO31, the half-life of cyclin D1 was substantially decreased (Supplementary Fig. 6).

The proteasome-mediated degradation of cyclin D1 requires phosphorylation on Thr 286 (refs 12–14). We therefore tested whether FBXO31-mediated degradation of cyclin D1 was dependent on Thr 286 phosphorylation. SK-MEL-28 cells were stably transfected with a plasmid expressing haemagglutinin (HA)-tagged cyclin D1 or a cyclin D1 derivative bearing a threonine-to-alanine substitution at position 286 (cyclin D1(T286A))[12] and then transduced with a retrovirus expressing FBXO31. Figure 2c shows that ectopic expression of FBXO31 resulted in the degradation of wild-type cyclin D1 but not the cyclin D1(T286A) mutant. After expression of cyclin D1(T286A), ectopically expressed FBXO31 failed to induce efficient G1 arrest (Fig. 2d and Supplementary Fig. 7a) or to block DNA synthesis (Fig. 2e).
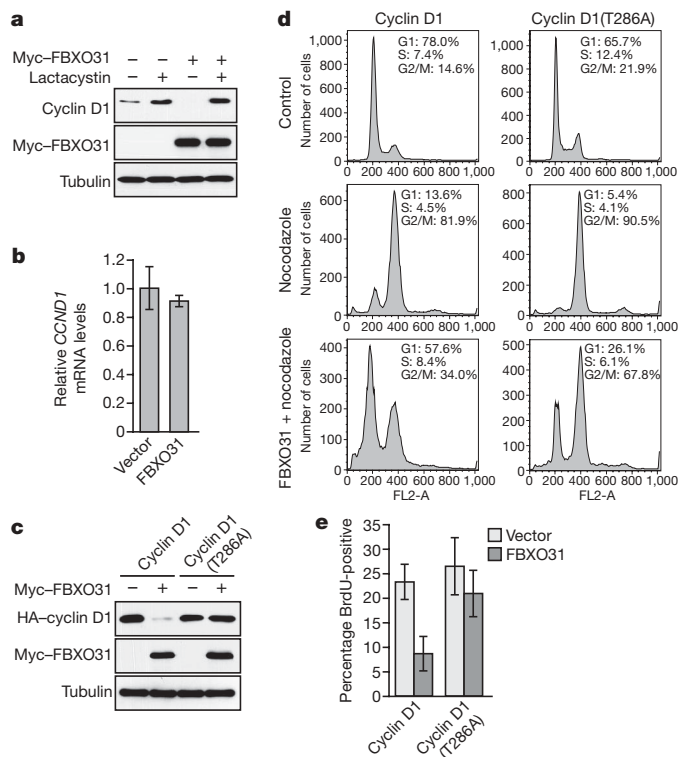


**Figure 2 | FBXO31-mediated cyclin D1 degradation occurs through the proteasomal pathway. a**, Cyclin D1 levels in SK-MEL-28 cells transduced with a retrovirus expressing FBXO31 or empty vector and treated in the presence or absence of lactacystin. **b**, Quantitative real-time RT–PCR monitoring of *CCND1* mRNA levels (error bars indicate s.d.; n = 3). **c**, Cyclin D1 levels in SK-MEL-28 cells stably transfected with a plasmid expressing either HA-tagged cyclin D1 or cyclin D1(T286A) and transduced with an FBXO31 retrovirus. **d**, FACS analysis of cells described in **c**. **e**, DNA replication assay (error bars indicate s.d.; n = 3).

To test whether FBXO31 and cyclin D1 interact, we performed co-immunoprecipitation experiments. SK-MEL-28 cells transduced with a retrovirus expressing Myc-tagged FBXO31 were stably transfected with a plasmid expressing HA-tagged cyclin D1 and treated with lactacystin to prevent cyclin D1 degradation (see Fig. 2a). The results of Fig. 3a show the presence of cyclin D1 in the FBXO31 immunoprecipitate, which was most evident when lactacystin was added. A similar result was obtained in the reciprocal co-immunoprecipitation. In contrast, FBXO31 failed to interact with cyclin D1(T286A) (Supplementary Fig. 7b). Figure 3a also shows, in agreement with previous reports[9], that FBXO31 was part of the SCF complex, as demonstrated by its association with SKP1 and CUL1. The co-immunoprecipitation experiment of Fig. 3b shows that there was also an interaction between the endogenous FBXO31 and cyclin D1 proteins.

To determine whether the F-box motif of FBXO31 was required for cyclin D1 proteolysis, we used an FBXO31 derivative in which the F-box had been deleted (FBXO31ΔF)[9]. In contrast with wild-type FBXO31, ectopic expression of FBXO31ΔF did not result in decreased levels of cyclin D1 (Fig. 3c). Consistent with this finding was our observation that ectopic expression of FBXO31, but not that of FBXO31ΔF, resulted in polyubiquitylation of cyclin D1 (Fig. 3d). As expected, cyclin D1(T286A) was not polyubiquitylated. Figure 3e shows that FBXO31 directed polyubiquitylation of cyclin D1 *in vitro* in the absence of any other F-box protein.

As noted above, it has previously been shown that oncogene-induced senescence involves activation of the DDR[1–3]. We therefore considered the possibility that in addition to its role in BRAF-induced senescence[4], FBXO31 might also be involved in mediating G1 arrest after DNA damage. Figure 4a shows that after induction of DNA damage by γ-irradiation, the levels of FBXO31 increased and, in agreement with previous reports[15], this was accompanied by a decrease in cyclin D1 levels. Quantitative real-time reverse-transcriptase-mediated polymerase chain reaction (RT–PCR) confirmed that FBXO31 and *CCND1* mRNA levels were unchanged after γ-irradiation (Supplementary Fig. 8). Consistent with this finding was our observation that in untreated SK-MEL-28 cells the low levels of FBXO31 resulted, at least in part, from proteasomal degradation (Supplementary Fig. 9). RNAi-mediated knockdown of FBXO31 prevented the decrease in cyclin D1 after γ-irradiation (Fig. 4b and Supplementary Fig. 10). In support of this conclusion, γ-irradiation decreased the half-life of cyclin D1 in SK-MEL-28 cells expressing a control non-silencing short hairpin RNA (shRNA) but not an FBXO31 shRNA (Supplementary Fig. 11). In FBXO31 knockdown (KD) SK-MEL-28 cells, γ-irradiation failed to induce G1 arrest (Fig. 4c and Supplementary Fig. 12).

Phosphorylation of cyclin D1 at Thr 286 can be mediated by glycogen synthase kinase 3β (GSK3β)[12,13] and through a mitogen-activated protein (MAP) kinase pathway[14]. We found that GSK3β was not required for cyclin D1 degradation after γ-irradiation, whereas blocking MAP kinase signalling with a chemical inhibitor of MAP kinase/ERK kinase (MEK) prevented cyclin D1 degradation after γ-irradiation (Supplementary Fig. 13).

ATM is a serine–threonine protein kinase with a critical role in the DDR[16–18]. In response to DNA double-strand breaks, ATM is auto-phosphorylated on several residues including Ser 1981, and functions by phosphorylating, and in many cases stabilizing, several downstream protein targets that function in the DDR[16,19]. FBXO31 contains two putative ATM phosphorylation sites at amino-acid residues 278/279 and 400/401, the former of which is conserved (Fig. 4d, top). Co-immunoprecipitation experiments revealed that FBXO31 and ATM interacted (Supplementary Fig. 14). A glutathione S-transferase (GST) fusion protein containing the 278/279 serine-glutamine site was phosphorylated by ATM *in vitro* (Fig. 4d, bottom), and an FBXO31 derivative bearing a serine-to-alanine mutation at position 278 failed to accumulate after γ-irradiation (Fig. 4e) or on co-transfection with ATM (Fig. 4f). In ATM KD SK-MEL-28 cells, FBXO31 was not induced after γ-irradiation and, accordingly, cyclin D1 levels remained unchanged (Fig. 4g and Supplementary Fig. 15).
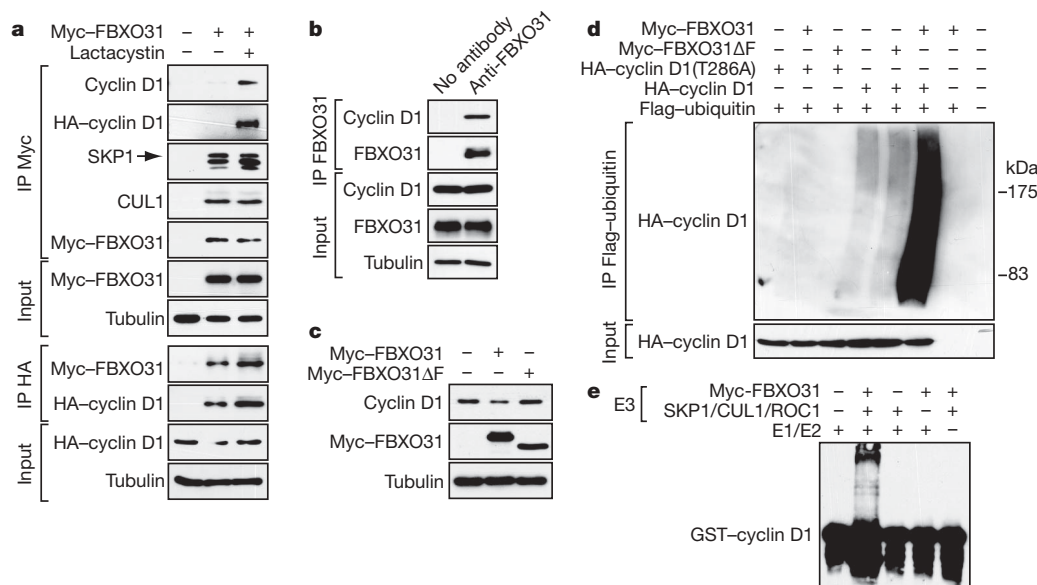
**Figure 3 | FBXO31 interacts with and directs ubiquitylation of cyclin D1.**
**a**, Co-immunoprecipitation of FBXO31 with cyclin D1 and SCF complex
components. The FBXO31-cyclin D1 interaction was specific
(Supplementary Fig. 18) and predominantly cytoplasmic (Supplementary
Fig. 19). IP, immunoprecipitation. **b**, Co-immunoprecipitation of
endogenous FBXO31 and cyclin D1. **c**, Cyclin D1 levels in SK-MEL-28 cells

ectopically expressing vector, FBXO31 or FBXO31ΔF. **d**, *In vivo*
ubiquitylation assay. Polyubiquitylated cyclin D1 was detected by
immunoprecipitation of Flag-tagged ubiquitin followed by immunoblotting
for HA–cyclin D1. **e**, *In vitro* ubiquitylation assay. Immunopurified SCF
complexes were incubated with GST–cyclin D1, Erk2, E1, E2, ATP and
ubiquitin.
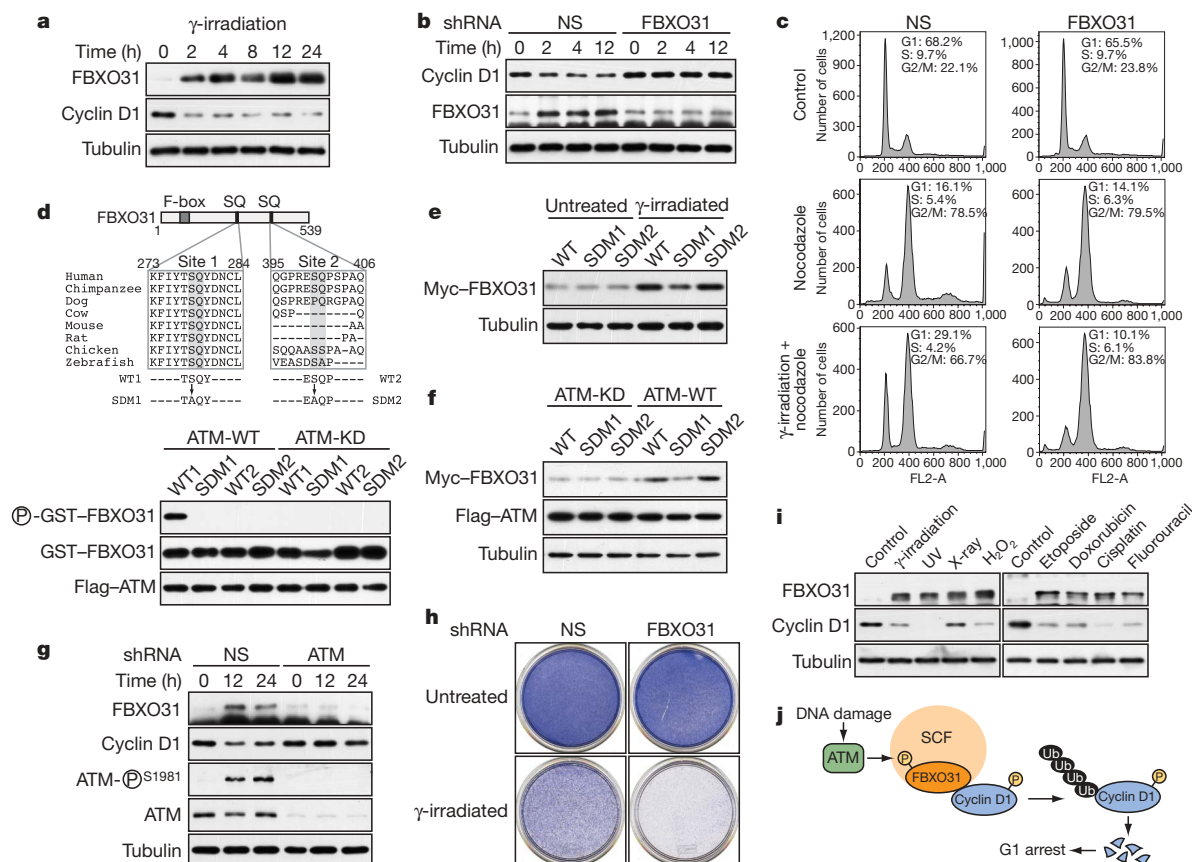


**Figure 4 | Cell cycle arrest after DNA damage requires ATM-mediated
induction of FBXO31. a**, Immunoblot analysis of FBXO31 after
γ-irradiation. For comparison with ectopically expressed FBXO31 levels see
Supplementary Fig. 20. **b**, Cyclin D1 levels in cells expressing a non-silencing
(NS) or FBXO31 shRNA after γ-irradiation. **c**, FACS analysis. **d**, Top:
putative ATM sites in FBXO31. Bottom: *in vitro* phosphorylation of
GST–FBXO31 fusion proteins harbouring wild-type (WT) or mutated

(SDM) SQ sites. ATM-WT, wild-type ATM; ATM-KD, kinase-dead ATM
mutant. **e**, **f**, Levels of ectopically expressed WT or mutant FBXO31 after
γ-irradiation (**e**) and ectopic ATM expression (**f**). **g**, FBXO31, cyclin D1,
ATM and phosphorylated ATM in γ-irradiated cells expressing a non-
silencing or ATM shRNA. **h**, Colony formation assay in untreated or
γ-irradiated cells. **i**, FBXO31 and cyclin D1 levels after treatment with DNA-
damaging agents. **j**, Model.

724

Knockdown of FBXO31 substantially increased the sensitivity of SK-MEL-28 cells to γ-irradiation (Fig. 4h). Taken together, these results indicate that after γ-irradiation ATM phosphorylates FBXO31 directly, which results in the stabilization of FBXO31 and the subsequent degradation of cyclin D1.

Finally, we tested whether other DNA-damaging agents also induced FBXO31. Figure 4i shows that FBXO31 levels greatly increased and cyclin D1 levels declined to varying extents after ultraviolet irradiation, X-ray irradiation, oxidative stress ($H_2O_2$) or addition of the chemotherapeutic DNA-damaging agents etoposide, adriamycin, cisplatin or fluorouracil. On the basis of these results we conclude that induction of FBXO31 is a general response to genotoxic stress.

Taken together, the results reported here show that FBXO31 is induced by DNA damage and, once stabilized, targets cyclin D1 for degradation by ubiquitin-mediated proteolysis, leading to G1 arrest (Fig. 4j). Our results address a long-standing problem in the DNA repair field. It has previously been shown that G1 arrest after DNA damage occurs in a two-phase response, referred to as initiation and maintenance[15]. The maintenance phase is a slow response that is due primarily to p53-dependent transcription induction of the cell cycle inhibitor p21. By contrast, initiation is a rapid, p53-independent response that results primarily from cyclin D1 degradation. However, the mechanistic basis by which cyclin D1 becomes rapidly degraded after genotoxic stress remained to be determined. Our results indicate that the initiation of G1 arrest after genotoxic stress is due to the induction of FBXO31, which then interacts with cyclin D1 and mediates its degradation.

Two other F-box proteins, FBXO4 (also called FBX4) and FBXW8, have previously been reported to mediate cyclin D1 degradation[14,20]. FBXO31, FBXO4 and FBXW8 probably function through a common mechanism, as demonstrated by the observations that all three proteins are incorporated into the SCF complex and require the phosphorylation of cyclin D1 at Thr 286 to direct cyclin D1 degradation. However, an important distinction between the three proteins is that FBXO4 and FBXW8 seem to be required for normal cell cycle progression[14,20], whereas we find no evidence for a cell-cycle defect or a decreased growth rate after knockdown of FBXO31 (ref. 4; see above and data not shown). Moreover, in contrast with FBXO31, the levels of FBXO4 and FBXW8 do not increase after γ-irradiation (Supplementary Fig. 16a), and knockdown of FBXO4 or FBXW8 does not affect cyclin D1 degradation in γ-irradiated SK-MEL-28 cells (Supplementary Fig. 16b). The results presented here reveal that FBXO31 is a dedicated checkpoint protein whose function is to arrest cells after genotoxic stress.

## METHODS SUMMARY

**Cell cycle analysis.** For FACS analysis, SK-MEL-28 cells transduced with retrovirus (multiplicity of infection 20) expressing FBXO31 or empty vector, in the presence or absence of a plasmid expressing cyclin D1 or cyclin D1(T286A), were collected 48 h after infection with the virus, after having been treated with 8 μg ml$^{-1}$ of nocodazole (Sigma) for the final 16 h. Cells were washed with cold PBS and fixed with ethanol. Treatment with nocodazole for γ-irradiation-treated cells was performed as described previously[15] after treatment of SK-MEL-28 cells with a non-silencing or FBXO31 shRNA. FACS samples were analysed with a FACSCalibur system (BD Biosciences).

For labelling with bromodeoxyuridine (BrdU), SK-MEL-28 cells transduced with retrovirus expressing FBXO31 or empty vector, in the presence or absence of a plasmid expressing cyclin D1 or cyclin D1(T286A), were collected 48 h after virus infection, after having been treated with BrdU (20 μM) for the final 4 h. Cells were then harvested, fixed in 70% ethanol and stained sequentially with an anti-BrdU mouse antibody (Calbiochem), fluorescein isothiocyanate-conjugated goat anti-mouse antibody (Calbiochem) and 4,6-diamidino-2-phenylindole (DAPI; Sigma). DAPI-stained cells were scored as BrdU-positive or BrdU-negative with a Zeiss Axiophot 2 fluorescence microscope; about 500–600 cells per sample were scored.

**Ubiquitylation assays.** For *in vivo* assays, SK-MEL-28 cells were co-transfected with plasmids expressing HA–cyclin D1 or HA–cyclin D1(T286A) and Flag-tagged ubiquitin and then transduced with an FBXO31 or FBXO31ΔF retrovirus. Polyubiquitylated cyclin D1 was detected by immunoprecipitation of ubiquitin with an anti-Flag antibody, followed by immunoblotting with an anti-HA antibody

for cyclin D1. For *in vitro* assays, 293T cells were transfected with vectors encoding Myc–FBXO31, Myc–CUL1, Myc–SKP1 and Myc–ROC1. The SCF$^{FBXO31}$ (E3) complexes were immunopurified from the cell lysate using an anti-Myc antibody and incubated with *in vitro*-translated GST-tagged cyclin D1 in the presence of purified, recombinant active Erk2, E1, E2 (UbcH5A), ATP and ubiquitin.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Bartkova, J. *et al.* Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature* **444**, 633–637 (2006).
2. Di Micco, R. *et al.* Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature* **444**, 638–642 (2006).
3. Mallette, F. A., Gaumont-Leclerc, M. F. & Ferbeyre, G. The DNA damage signaling pathway is a critical mediator of oncogene-induced senescence. *Genes Dev.* **21**, 43–48 (2007).
4. Wajapeyee, N., Serra, R. W., Zhu, X., Mahalingam, M. & Green, M. R. Oncogenic BRAF induces senescence and apoptosis through pathways mediated by the secreted protein IGFBP7. *Cell* **132**, 363–374 (2008).
5. Lin, Y. W. *et al.* Deletion mapping of chromosome 16q24 in hepatocellular carcinoma in Taiwan and mutational analysis of the 17β-HSD gene localized to the region. *Int. J. Cancer* **93**, 74–79 (2001).
6. Launonen, V. *et al.* Loss of heterozygosity at chromosomes 3, 6, 8, 11, 16, and 17 in ovarian cancer: correlation to clinicopathological variables. *Cancer Genet. Cytogenet.* **122**, 49–54 (2000).
7. Harkonen, P., Kyllonen, A. P., Nordling, S. & Vihko, P. Loss of heterozygosity in chromosomal region 16q24.3 associated with progression of prostate cancer. *Prostate* **62**, 267–274 (2005).
8. Miller, B. J., Wang, D., Krahe, R. & Wright, F. A. Pooled analysis of loss of heterozygosity in breast cancer: a genome scan provides comparative evidence for multiple tumor suppressors and identifies novel candidate regions. *Am. J. Hum. Genet.* **73**, 748–767 (2003).
9. Kumar, R. *et al.* FBXO31 is the chromosome 16q24.3 senescence gene, a candidate breast tumor suppressor, and a component of an SCF complex. *Cancer Res.* **65**, 11304–11313 (2005).
10. Ho, M. S., Tsai, P. I. & Chien, C. T. F-box proteins: the key to protein degradation. *J. Biomed. Sci.* **13**, 181–191 (2006).
11. Alao, J. P. The regulation of cyclin D1 degradation: roles in cancer development and the potential for therapeutic invention. *Mol. Cancer* **6**, 24 (2007).
12. Diehl, J. A., Zindy, F. & Sherr, C. J. Inhibition of cyclin D1 phosphorylation on threonine-286 prevents its rapid degradation via the ubiquitin–proteasome pathway. *Genes Dev.* **11**, 957–972 (1997).
13. Diehl, J. A., Cheng, M., Roussel, M. F. & Sherr, C. J. Glycogen synthase kinase-3β regulates cyclin D1 proteolysis and subcellular localization. *Genes Dev.* **12**, 3499–3511 (1998).
14. Okabe, H. *et al.* A critical role for FBXW8 and MAPK in cyclin D1 degradation and cancer cell proliferation. *PLoS ONE* **1**, e128 (2006).
15. Agami, R. & Bernards, R. Distinct initiation and maintenance mechanisms cooperate to induce G1 cell cycle arrest in response to DNA damage. *Cell* **102**, 55–66 (2000).
16. Shiloh, Y. The ATM-mediated DNA-damage response: taking shape. *Trends Biochem. Sci.* **31**, 402–410 (2006).
17. Elledge, S. J. Cell cycle checkpoints: preventing an identity crisis. *Science* **274**, 1664–1672 (1996).
18. Zhou, B. B. & Elledge, S. J. The DNA damage response: putting checkpoints in perspective. *Nature* **408**, 433–439 (2000).
19. Matsuoka, S. *et al.* ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**, 1160–1166 (2007).
20. Lin, D. I. *et al.* Phosphorylation-dependent ubiquitination of cyclin D1 by the SCF$^{FBX4-αB\ crystallin}$ complex. *Mol. Cell* **24**, 355–366 (2006).

**Author Contributions** M.K.S. and N.W. contributed equally to this work. M.K.S., N.W. and M.R.G. conceived and designed the experiments. M.K.S. and N.W. performed the experiments and, together with M.R.G., analysed the data and wrote the paper.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.R.G. (michael.green@umassmed.edu).

725

# METHODS

**Cell lines and culture.** Human melanoma SK-MEL-28 cells were obtained from the American Type Culture Collection and grown as recommended. SK-MEL-28 cell lines expressing Myc–FBXO31 or Myc–FBXO31ΔF were generated by retroviral transduction. In brief, SK-MEL-28 cells were seeded ($10^6$ cells per 10-cm dish) and on the next day they were transduced with 10 ml of FBXO31-expressing or FBXO31ΔF-expressing retrovirus (multiplicity of infection 20) using $10 \, \mu g \, ml^{-1}$ Polybrene (Sigma). The medium was changed 12 h after virus infection.

**Plasmids.** Retroviral vectors expressing Myc–FBXO31 and Myc–FBXO31ΔF were described previously[9] and provided by R. Kumar. Plasmids expressing HA–cyclin D1 and HA–cyclin D1(T286A) were described previously[21] and purchased from Addgene (plasmids 11181 and 11182, respectively). Other plasmids were obtained as follows: Flag–ubiquitin (S. Cook), Flag–FBXW6 (M. Pagano), V5-FBXW8 (O. Tetsu) and wild-type and kinase-dead Flag-ATM (M. Kastan).

**Immunoblot analysis.** Protein extracts were prepared by lysis in a buffer containing (50 mM Tris-HCl pH 7.4, 0.1% Triton X-100, 5 mM EDTA, 250 mM NaCl, 50 mM NaF, 0.1 mM Na$_3$VO$_4$) and protease inhibitors in the presence or absence of a phosphatase inhibitor cocktail (Sigma), as needed. Blots were probed with the following antibodies: anti-human cyclin D1 mouse monoclonal (BD Bioscience Pharmingen), anti-phospho cyclin D1(T286) rabbit polyclonal (Cell Signaling), anti-cyclin D2 mouse monoclonal (Abcam), anti-cyclin D3 mouse monoclonal (Santa Cruz), anti-cyclin B1 mouse monoclonal (Upstate), anti-cyclin E2 rabbit monoclonal (Abcam); anti-CDK1 mouse monoclonal (BD Bioscience Pharmingen), anti-CDK2 rabbit polyclonal (Abcam), anti-CDK4 rabbit polyclonal (Santa Cruz), anti-CDK6 rabbit polyclonal (Santa Cruz), anti-p14 rabbit polyclonal (Santa Cruz), anti-p16 rabbit polyclonal (Cell Signaling), anti-p27 mouse monoclonal (Santa Cruz), anti-p57 mouse monoclonal (Sigma), anti-tubulin mouse monoclonal (Sigma), anti-HA mouse monoclonal (Roche), anti-myc mouse monoclonal (Roche), anti-Flag mouse monoclonal (Sigma), anti-SKP1 mouse monoclonal (BD Bioscience Pharmingen), anti-CUL1 rabbit polyclonal (Cell Signaling), anti-actin mouse monoclonal (Sigma), anti-ATM rabbit monoclonal (Cell Signaling), anti-ATM-P(S1981) mouse monoclonal (Cell Signaling), anti-GSK3β mouse monoclonal (BD Transduction Laboratories), anti-p44/42 (Erk1/2) rabbit polyclonal (Cell Signaling), anti-phospho-p44/42 (Erk1/2) rabbit monoclonal (Cell Signaling), anti-FBXO4 rabbit monoclonal (Rockland Immunochemicals), anti-FBXW8 rabbit polyclonal (Saier Biotechnology Inc.) and anti-PARP1 mouse monoclonal (EMD Biosciences, Inc.). To raise the rabbit polyclonal anti-FBXO31 antibody, a mixture of two peptides (EGRGRQGPRESQPSPAQPRAEC and CPVGVS SRNEDYPRT) was used to immunize rabbits (AnaSpec) followed by affinity purification. The antibody was used to detect endogenous FBXO31 by immunoprecipitation–immunoblot analysis. Where indicated, 5 μM lactacystin (Calbiochem) was added for 8 h before preparation of the protein extract.

**Quantitative real-time RT–PCR.** Total RNA was isolated and reverse transcription was performed as described[22], followed by quantitative real-time PCR with Platinum SYBR Green qPCR SuperMix-UDG with Rox (Invitrogen). Primer sequences were as follows: cyclin D1, 5′-TCTACACCGACAACTCCATCC-3′ (forward) and 5′-TTCCACTTGAGCTTGTTCACC-3′ (reverse); FBXO31, 5′-AATCCGGCCTTTTGACCAGA-3′ (forward) and 5′-TCCGCTCACAGGA AGAGCAC-3′ (reverse).

**RNA interference.** FBXO31 and ATM knockdown cell lines were generated by stable transduction of $6 \times 10^4$ SK-MEL-28 cells with the following shRNAs from Open Biosystems: FBXO31 (source location V2HS_157523, clone location SH2232-d-1 and source location V2HS_25023, clone location SH2497-b-9), ATM (source location V2HS-89368, clone location SH2178-a-12 and source location V2HS-192800, clone location SH116-a-9), FBXO4 (source location V2HS-15758, clone location SH2245-b-3), GSK3β (source location V2HS-114292, clone location SH2049-a-7) and non-silencing (5′-TCTCGCTTGGGC GAGAGTAAG-3′). siRNAs were purchased from Sigma; their sequences (sense strand) were as follows: FBXO31, 5′-CUGAUGAAGUUCAUCUACAUU-3′; FBXW8, 5′-AAGAUGUGCACAGGUGAGCAA-3′; FBXO4, 5′-CAAUGAAG GUGAUGAUCAAUU-3′; cyclin D1, 5′-CUGUCCCACUCCUACGAUAUU-3′ and 5′-CUGUGUAUCUCUUUCACAUUU-3′; luciferase, 5′-UGAUCAAAUAC AAGGGAUA-3′.

**Generation of FBXO31 mutants.** The plasmid pLNCX2-FBXO31 (ref. 9) was used as a template to generate FBXO31 mutants with the QuikChange XL site-directed mutagenesis kit (Stratagene). To generate in-frame GST-tagged fusion proteins, the wild-type and mutant FBXO31 sequences were amplified by PCR (with primers WT1_forward (5′-CGGGATCCGCTGCCACAGTGGAGTGC-3′) and WT1_reverse (5′-GGAATTCGATGAGGTCGTCGGGGCG-3′) or WT2_forward (5′-CGGAATTCCGGATCCAGCTGCCCGAC-3′) and WT2_reverse (5′-ACGTCGACCACGGGCAGCACG AACGG-3′)), digested with BamH1 and EcoR1, and cloned into pGEX-4T-1 (Amersham). All constructs were confirmed by sequencing.

**ATM kinase assays.** *In vitro* ATM kinase assays were performed as described previously[23]. In brief, 293T cells were transfected with plasmids (10 μg) expressing Flag-tagged ATM or the kinase-dead mutant (ATM-KD). Cell lysates were prepared by resuspension in modified TGN buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% Tween 20, 0.3% Nonidet P40, 1 mM sodium fluoride, 1 mM Na$_3$VO$_4$, 1 mM phenylmethylsulphonyl fluoride, and a protease inhibitor mixture from Roche Molecular Biochemicals). Cell lysates were incubated with anti-Flag M2 antibody (Sigma) and Protein A/G-agarose overnight at 4 °C. The precipitated beads were washed with TGN buffer followed by TGN buffer plus 0.5 M LiCl, and two washes with kinase buffer (20 mM HEPES pH 7.5, 50 mM NaCl, 10 mM MgCl$_2$, 1 mM dithiothreitol, 10 mM MnCl$_2$). Finally, the pellet was resuspended in 50 μl of kinase buffer containing 10 μCi of [γ-$^{32}$P]ATP and 1 μg of GST fusion substrate (described above). The kinase reaction was incubated for 20 min at 30 °C and stopped by the addition of SDS–PAGE protein sample buffer. Proteins were separated by SDS–PAGE, transferred to nitrocellulose and detected by autoradiography.

**Colony formation assays.** SK-MEL-28 cells ($5 \times 10^5$) stably expressing a non-silencing or FBXO31 shRNA were plated in 60-mm dishes and γ-irradiated at a dose of 10 Gy. Cells were selected with puromycin for 14 days, then stained with crystal violet and photographed.

**Induction of DNA damage.** SK-MEL-28 cells were grown as a subconfluent monolayer and treated with either γ-ray (20 Gy), ultraviolet (50 mJ cm$^{-2}$) or X-ray (20 Gy) irradiation and then incubated at 37 °C before preparation of the protein extract. For H$_2$O$_2$ treatment, cells were treated with 0.02% H$_2$O$_2$ for 4 h. For chemotherapeutic DNA-damaging agents, cells were treated with 2 μM etoposide (Sigma), 0.2 μg ml$^{-1}$ doxorubicin (Sigma), 2 μg ml$^{-1}$ cisplatin (Sigma), 50 μM fluorouracil or, as a control, dimethylsulphoxide for 48 h before preparation of the protein extract.

**Cell cycle analysis by two-colour FACS.** Cells transduced with a retrovirus expressing FBXO31 or empty vector in the presence or absence of plasmids expressing cyclin D1 or cyclin D1(T286A), or cells stably expressing luciferase or cyclin D1 siRNAs, were incubated with BrdU (20 μM) for 4 h, fixed in 70% ethanol and stained sequentially with an anti-BrdU mouse antibody (Calbiochem), Alexa-488-conjugated goat anti-mouse antibody (Invitrogen) and propidium iodide (Sigma). FACS samples were analysed with a FACSCalibur system (BD Biosciences).

21. Newman, R. M. *et al.* Antizyme targets cyclin D1 for degradation. A novel mechanism for cell growth repression. *J. Biol. Chem.* **279**, 41504–41511 (2004).
22. Gazin, C., Wajapeyee, N., Gobeil, S., Virabasius, C. M. & Green, M. R. An elaborate pathway required for Ras-mediated epigenetic silencing. *Nature* **449**, 1073–1077 (2007).
23. Canman, C. E. *et al.* Activation of the ATM kinase by ionizing radiation and phosphorylation of p53. *Science* **281**, 1677–1679 (1998).

# LETTERS

# The structure of a cytolytic α-helical toxin pore reveals its assembly mechanism

Marcus Mueller[1]†, Ulla Grauschopf[1]†, Timm Maier[1], Rudi Glockshuber[1] & Nenad Ban[1]

Pore-forming toxins (PFTs) are a class of potent virulence factors that convert from a soluble form to a membrane-integrated pore[1]. They exhibit their toxic effect either by destruction of the membrane permeability barrier or by delivery of toxic components through the pores. Among the group of bacterial PFTs are some of the most dangerous toxins, such as diphtheria and anthrax toxin. Examples of eukaryotic PFTs are perforin and the membrane-attack complex, proteins of the immune system[2]. PFTs can be subdivided into two classes, α-PFTs and β-PFTs, depending on the suspected mode of membrane integration, either by α-helical or β-sheet elements[3]. The only high-resolution structure of a transmembrane PFT pore is available for a β-PFT—α-haemolysin from *Staphylococcus aureus*[4]. Cytolysin A (ClyA, also known as HlyE), an α-PFT, is a cytolytic α-helical toxin responsible for the haemolytic phenotype of several *Escherichia coli* and *Salmonella enterica* strains[5–8]. ClyA is cytotoxic towards cultured mammalian cells, induces apoptosis of macrophages and promotes tissue pervasion[9–11]. Electron microscopic reconstructions demonstrated that the soluble monomer of ClyA[12] must undergo large conformational changes to form the transmembrane pore[13,14]. Here we report the 3.3 Å crystal structure of the 400 kDa dodecameric transmembrane pore formed by ClyA. The tertiary structure of ClyA protomers in the pore is substantially different from that in the soluble monomer. The conversion involves more than half of all residues. It results in large rearrangements, up to 140 Å, of parts of the monomer, reorganization of the hydrophobic core, and transitions of β-sheets and loop regions to α-helices. The large extent of interdependent conformational changes indicates a sequential mechanism for membrane insertion and pore formation.

The assembled ClyA pore forms a hollow cylinder with a height of 130 Å, an inner diameter of 70 Å and a maximum outer diameter of 105 Å (Fig. 1). The top is constricted to outer and inner diameters of ~90 Å and ~35 Å, in good agreement with the 30 Å diameter pore opening estimated from osmotic protection assays[9,15]. Protomers are arranged in a parallel head-to-head orientation. Each protomer is organized as a three-helix bundle with the three helices (αB, αC and αF) running along the wall of the pore (Figs 1 and 2b, d). At the top of this bundle the amino-terminal helix (αA) of each protomer is protruding into an iris-like arrangement forming the constriction at the top of the pore. On the opposite end, the carboxy-terminal helix (αG) is flanking the three-helix bundle on the outside of the pore. This structure of the protomer is substantially different from that of the previously determined 34 kDa ClyA soluble monomer, which is an elongated, mostly α-helical molecule[12] (Fig. 2a, c) with a core bundle of four long α-helices (αA, αB, αC and αF). A hydrophobic β-hairpin, the 'β-tongue', which is flanked by two short helices and packs against the four helix bundle, builds the 'head' domain of the molecule. The β-tongue is mostly hydrophobic and is located

in a predicted transmembrane region, encompassing residues 177 to 203 (ref. 5; Supplementary Fig. 1).

In the assembled pore all helices, except αG, contribute to protomer–protomer contacts. Each helix of the twisted three-helix bundle contributes to both interfaces (Supplementary Fig. 2a, b). Each interface comprises an area of ~2,400 Å², with more than one-quarter of a protomer's surface buried in the two interfaces. A network of 25 hydrogen bonds and 13 salt bridges indicates tight and specific association of protomers (Supplementary Fig. 2c). These specific interfaces do not exist in the soluble ClyA monomer, indicating that conversion of ClyA to the protomeric form has to precede oligomerization.

An analysis of the electrostatic potential on the surface of the pore shows a 30-Å-wide hydrophobic stretch at the top, indicating the membrane-inserted region (Fig. 1e). The location of this stretch coincides with rim-like electron density features in experimental low-resolution maps, probably originating from weakly ordered maltose moieties of the detergent dodecylmaltoside (DDM) bound to this region (Supplementary Fig. 4). This hydrophobic portion encompasses residues 177 to 203 and is organized as a helix–turn–helix motif, projecting 20 Å from a rim of charged residues—a distance not sufficient to span a membrane. The remaining stretch of hydrophobic surface that crosses the membrane is built from the apolar residues of the amphipathic helix A1, whereas the charged residues line the inside of the pore's constriction. The inner surface of the pore has a strong negative potential (Fig. 1f), a feature reflected in the cation selectivity of the pore[15].

On pore formation, ClyA converts from a four- to a three-helix bundle as a result of a large translocation of αA. In the monomer, the N-terminal helix is embedded in a groove between αF and αG, forming a part of the five-helix bundle at the tail of the molecule. A distinct kink at residue Pro 36 divides αA in two parts, αA1 and αA2. In the protomer, αA is turned upwards, resulting in repositioning of its N terminus by more than 140 Å. Displacement of αA causes repacking of the hydrophobic core of the lower half of the helix bundle, positioning αF into the space occupied by αA in the monomer. Helix A2 together with the loop and the short 3₁₀ helix connecting it to αB refold into an α-helical extension of αB. The N-terminal end of αA forms the constriction at the top of the pore, protruding from the three-helix bundle of the protomer. Pro 36 now induces a kink between helices αA and αB of the protomer (Fig. 3b).

As a result of the complete rearrangement of the head region, the protomer is devoid of any β-sheets (Fig. 3b, d). In the monomer, this head region is composed of the β-tongue, flanked by αD and αE. In the protomer, αD together with the connecting loop turns into an extension of αC. A loop at the end of the extended αC is partially composed of residues that belong to the β-tongue in the monomer. The remaining residues of the β-tongue are transformed into an α-helical extension of αE. This β-strand-to-α-helix conversion contrasts the

---
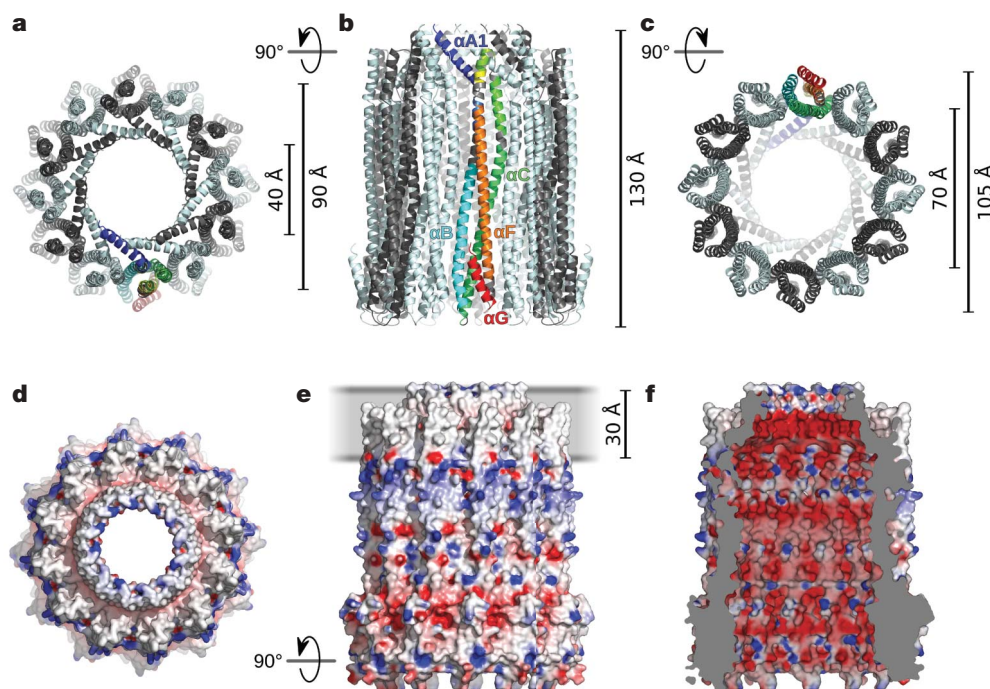
**Figure 1 | Structure of the ClyA pore.** **a–c**, Top (**a**), side (**b**) and bottom (**c**) view in ribbon representation. One protomer is highlighted, with the secondary structure elements coloured from blue to red from N to C terminus; other protomers are shown alternating in pale blue and grey. The protomers are oriented head-to-head and parallel to the symmetry axis of the pore. **d–f**, Surface representation with mapped electrostatic potential from negative (red) to positive (blue) in top (**d**), side (**e**) and inside (**f**) view. **e**, The membrane-inserted part of the pore is easily recognized as a 30-Å-wide apolar region at the top of the molecule. **f**, The inside of the pore shows a strong negative potential on most of its surface, especially in the upper constriction of the pore.

transformation of α-helices to β-hairpins observed for the β-PFT family of cholesterol-dependent cytolysins[16].

Most of the loop between αF and αG in the monomer transforms into a helical extension of αF, whereas the first four residues of αG become part of the connecting loop in the protomer. Only small conformational changes occur for helices αB and αC, which can be considered as a platform for the rearrangements outlined above.

Overall, 167 residues, more than half of the ClyA polypeptide, are affected by the topological reorganization of the protein, and 48 of these residues belong to a different type of secondary structure element in the two forms of the protein. The content of α-helices increases from 79% to 83%, in agreement with circular dichroism spectroscopy data[13].

The β-tongue and its surrounding residues have an important role in membrane association and triggering of the conformational switch, based on the following observations. First, the β-tongue is nearly exclusively composed of hydrophobic residues that are partially solvent-exposed. Insertion into a lipid bilayer is energetically favourable because it provides a hydrophobic environment for all residues. Second, the 'swinging-out' movement of the β-tongue on insertion is facilitated by a hinge formed by conserved glycines Gly 180 and Gly 201 flanking this region (Fig. 3). In fact, replacing the conserved Gly 180 by valine compromises the haemolytic activity of ClyA[17]. Third, the amino acid sequence of the β-tongue, with its high percentage of glycines and alanines, several bulky hydrophobic residues and a conserved central proline, has characteristics similar to internal fusion peptides of viral fusion proteins[18] and was predicted to be part of a transmembrane helix[5].

The conformational changes initiated by the swinging-out of the β-tongue trigger further molecular events. In the monomer, a phenylalanine at the tip of the β-tongue (Phe 190) is located at the centre of a cluster of five aromatic residues, forming a network of hydrophobic interactions (Fig. 3a, c). Of these, residues Phe 50 and Tyr 54 are located in the loop between αA1 and αB, and Phe 159 and Tyr 165

are located at the edges of the loop connecting αC and αD. These two pairs of aromatic residues appear to stabilize the loops that interrupt the helices. Upon swinging out of the central Phe 190, the hydrophobic core of the head region is disrupted and the loops where the two pairs of aromatic residues reside are energetically destabilized, leading to the formation of continuous helices (Fig. 3b, d).

According to this mechanistic model the two loops that become destabilized can be considered as springs, whereas the glycines flanking the β-tongue, and the phenylalanine at its tip, function as a hinge and a latch that locks the double-spring-loaded mechanism. Removing the latch triggers the springs and initiates the transformation of the soluble form of ClyA into its assembly-competent state.

Considering that many of the described conformational changes are interdependent and must occur sequentially, we propose a model for the series of conformational changes resulting in membrane insertion and pore assembly (Fig. 4). The membrane association of ClyA is facilitated by the hydrophobic character of exposed portions of the β-tongue (Fig. 4a). It may also include specific interactions, because ClyA shows different specificity towards membranes from different species, although no specific lipidic or proteinaceous receptors have been described so far[6,19]. Once associated, spontaneous insertion of the β-tongue is favoured owing to its strong hydrophobic character (Fig. 4b). The aromatic residue at its tip (Phe 190) will swing out of the aromatic cluster and unlatch the two springs. One of the springs involves a relatively small change in the angle between flanking helices and is likely to straighten first followed by a 180° swing of αA away from the helix αB/C platform (Fig. 4c). This leaves a binding site on the platform available for αF, which shifts upwards by two helical turns and positions itself in a slot previously occupied by αA (Fig. 4d). The upwards shift of αF occurs concurrently with the formation of a membrane helix from the portions of the membrane-inserted β-tongue (Fig. 4e). At this stage, the membrane-inserted helix–turn–helix motif is completed and the conformation of the protomer is very close to final. The oligomerization of the protomer can start although

no portion of the toxin has perforated the membrane yet. This is in contrast to some β-PFTs, which require no substantial conformational changes to form pre-pores[20,21]. (Supplementary Fig. 2d, e). The amphipathic αA1 is at this stage laying on the surface of the membrane as it is

flexibly disposed to the rest of the protomer due to the presence of the helix-breaking Pro 36, which serves as a hinge (Fig. 3b). Closing of the pore will lead to compaction and buckling of the amphipathic helices. As a result, these helices will be pushed upwards wedging, rather then 'punching', a hole in the membrane (Fig. 4f, g).

The energetic cost of the final step of membrane penetration is offset by several factors. The insertion of the helix–turn–helix motif and its
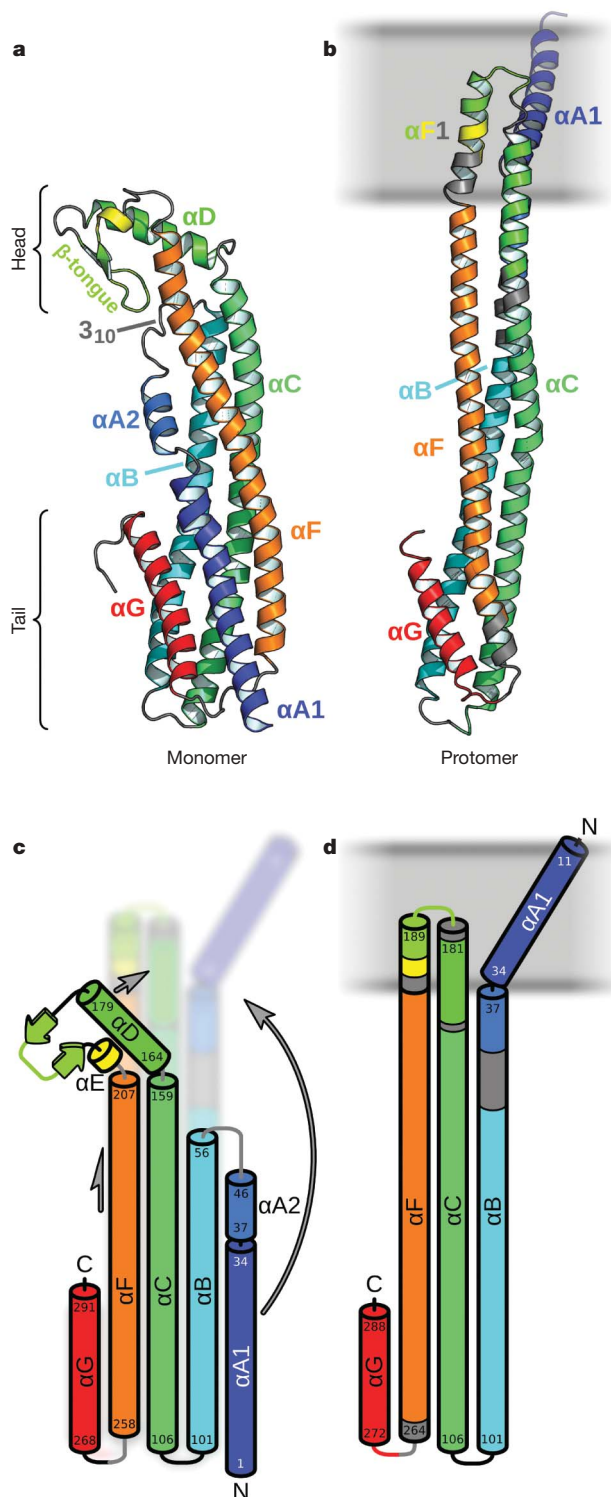


**Figure 2 | Conformational differences between the soluble and the protomeric form.** The monomer (**a**, **c**) and protomer (**b**, **d**) are shown in ribbon (**a**, **b**) and schematic (**c**, **d**) representation. The secondary structure elements are highlighted, each in a separate colour, from the N (blue) to the C (red) terminus. Loop regions of the monomer, which become part of an α-helix in the protomer, are coloured grey. The extent of the secondary structure elements is indicated by residue numbers in the schematic representations.
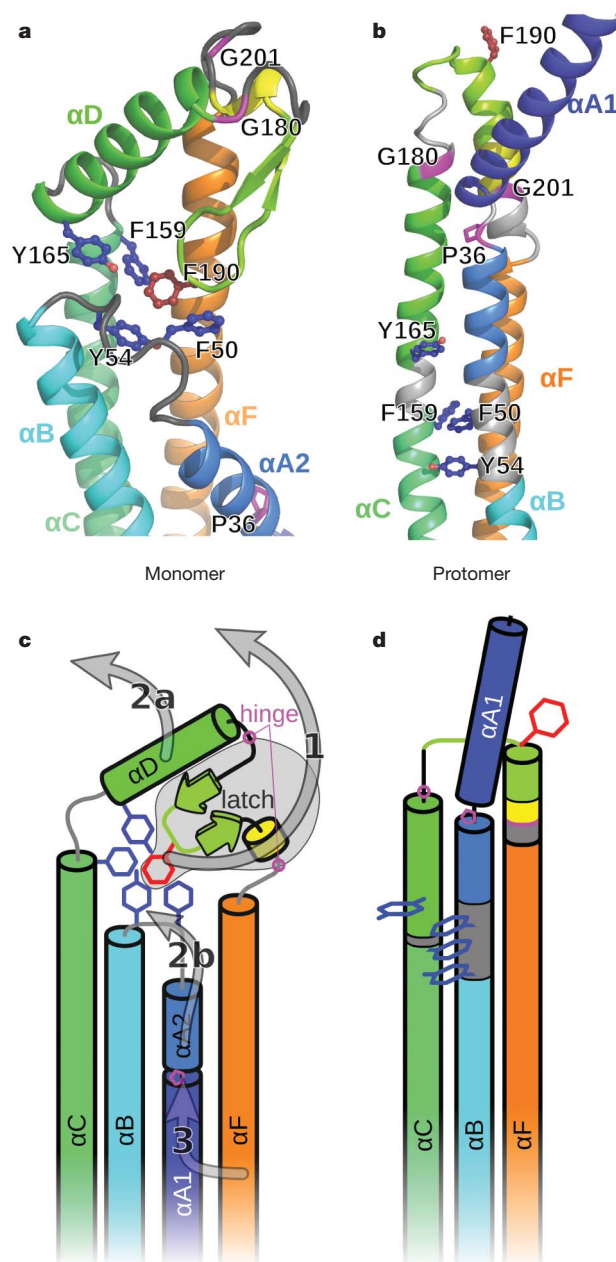
**Figure 3 | Key residues for triggering the conformational switch. a**, A cluster of five aromatic residues (Phe 50, Tyr 54, Phe 159, Tyr 165 and Phe 190, shown as ball-and-stick) is located at the tip of the β-tongue in the head region of the monomer. Two glycines at the top of the head region (Gly 180 and Gly 201, highlighted in magenta) facilitate swinging-out of the β-tongue upon membrane association. This removes the central Phe 190 (red) from the cluster of aromatic residues and destabilizes the loops neighbouring the cluster. **b**, In the protomer, the loop regions around the aromatic cluster of the monomer have turned into parts of joint helices. Pro 36, inducing a kink between helices αA1 and αB, is shown by stick representation in magenta. **c**, Schematic representation of **a** with the sequence of conformational changes indicated. Pro 36 is shown as a magenta pentagon. **d**, Schematic representation of **b**.
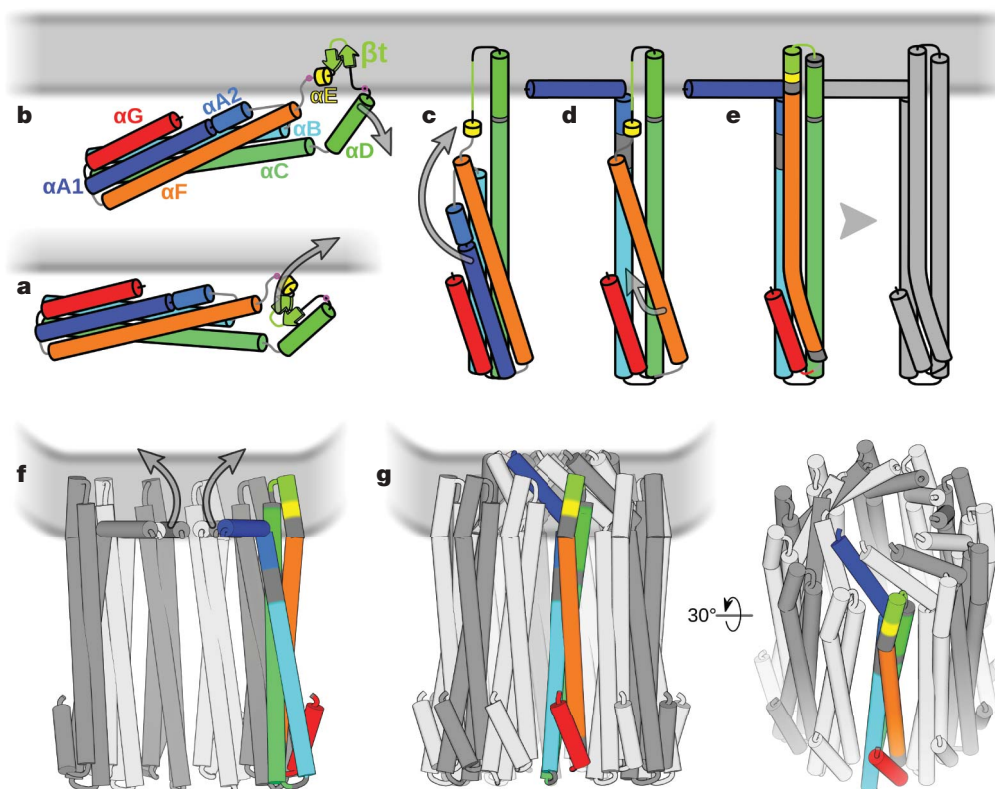
**Figure 4 | Mechanism of pore formation. a,** Swinging out of the β-tongue (βt, green–yellow) upon membrane association triggers the structural transformation. **b,** Helix D (green) and its connecting loop turn into a helical extension of αC (light green). **c,** Helix A (dark-blue and blue) departs from its position between αF and αG and swings upwards. Helix A2 and its connecting residues turn into an extension of αB (turquoise). **d,** The amphipathic helix A1 (dark blue) attaches to the membrane. Helix F (orange) moves into the position of αA1 next to αG (red), and rearrangements between αF and αG

occur. **e,** The repositioning of αF allows αF1 to form from the residues that were previously part of the monomer's head region. The membrane-inserted helix–turn–helix motif folds. The protomer interfaces have been generated and oligomerization can start. **f,** Progressing oligomerization leads to distortions of the membrane due to the dense packing of αA1 and the helix–turn–helix motif that spans only a portion of the bilayer. Helices A1 are wedged through the destabilized membrane. **g,** Iris-like, interlocking arrangement of helices A1 stabilizes the opening of the pore.

accumulation upon oligomerization will destabilize the membrane by inducing curvature in the lipid bilayer[22,23]. The accumulation of the amphipathic helices on the inside of the forming pore in the outer leaflet of the membrane will have similar effects[23]. The formation of extensive inter-subunit contacts during pore formation will stabilize upward tilting of the amphipathic A1 helices. Once formed, the pore opening is maintained by the iris-like interlocking of the amphipathic helices (Fig. 4g). After membrane penetration the amphipathic helices are spanning the bilayer and the curvature-inducing effects should be decreased.

Experimental evidence supports our model of pore formation. Membrane deformation by the partially inserted helix–turn–helix motif has been observed in electron micrographs of membranes and lipid vesicles exposed to ClyA[12–14,24]. Biochemical and spectroscopic analysis has shown that pore formation occurs in three kinetically distinct steps: fast association with membrane or detergent and slow conversion of the monomer followed by fast oligomerization[13]. This agrees with our model, in which tight association with the membrane can readily occur, whereas the complex conversion to the protomer conformation is a slow process.

The conformational differences between the soluble and protomeric form of ClyA are more extensive than those observed for α-haemolysin[4,25]. The extent of conformational changes observed in ClyA also exceeds the rearrangements anticipated in all previous models of ClyA assembly[12,13,26]. The mechanism of pore formation and the architecture of the membrane pore described here for ClyA may also apply to a pore-forming toxin from *Bacillus cereus* (B component of haemolysin BL) owing to a related overall architecture of the monomer in spite of very low sequence homology (Supplementary Figs 5 and 6).

The structural transformations observed during ClyA pore formation resemble the spring-loaded mechanism of influenza haemagglutinin and other type I viral fusion proteins[27,28]. In both mechanisms, loop-separated helices transform into a three-helix bundle of elongated helices (Supplementary Fig. 7). Furthermore, the sequence of the β-tongue has characteristics of an internal fusion peptide of a viral fusion protein[18].

The series of interdependent conformational changes by ClyA provides unique opportunities to study the process of membrane protein folding, spontaneous membrane insertion and pore formation. It may also be possible to use this molecule as a model system to develop predictive rules that will aid in understanding of molecular events that govern related cellular processes such as membrane fusion of cellular compartments and viral membrane fusion.

**METHODS SUMMARY**

ClyA from *E. coli* K12 was recombinantly expressed and purified as described previously[13]. Pores were assembled by incubating ClyA ($0.1\,\mathrm{mg\,ml^{-1}}$) in Dulbecco's phosphate-buffered saline (DPBS) and 2 mM dithiothreitol (DTT) in the presence of 0.1% DDM at 4 °C for 12–16 h. Crystals were obtained by sitting-drop vapour diffusion against 100 mM Tris-HCl (pH 8.4), 100–160 mM $MgCl_2$ and 14–18% PEG400. Diffraction properties of the severely anisotropically diffracting crystals were substantially improved by derivatization with sodium ethylmercurithiosalicylate.

Data were collected at beamline X06SA of the SLS (PSI, Villigen, Switzerland). The oligomeric state and symmetry of the crystallized pore were analysed with self-rotation functions (Supplementary Fig. 8). The Hg-substructure was determined in single-wavelength anomalous diffraction (SAD) data of the monoclinic crystal form (Supplementary Table 1). Initial phases were improved by density modification and a preliminary model was built into the 6 Å map of the mono-

clinic crystal form. The preliminary model was used for molecular replacement in data of the triclinic crystal form (Supplementary Table 1). Phases were extended from 6 Å to 3.3 Å by density modification. The model was built in COOT[29] and refined with PHENIX.refine[30] with six translation/libration/screw (TLS) and non-crystallographic symmetry (NCS) groups per protomer in the final cycles of refinement. The final model, comprising residues 8–292, agrees with the crystallographic data with $R_{free}$ and $R_{work}$ of 24.4% and 22.8%, respectively.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Parker, M. W. & Feil, S. C. Pore-forming protein toxins: from structure to function. *Prog. Biophys. Mol. Biol.* **88,** 91–142 (2005).
2. Voskoboinik, I., Smyth, M. J. & Trapani, J. A. Perforin-mediated target-cell death and immune homeostasis. *Nature Rev. Immunol.* **6,** 940–952 (2006).
3. Gouaux, E. Channel-forming toxins: tales of transformation. *Curr. Opin. Struct. Biol.* **7,** 566–573 (1997).
4. Song, L. Z. *et al.* Structure of staphylococcal α-hemolysin, a heptameric transmembrane pore. *Science* **274,** 1859–1866 (1996).
5. del Castillo, F. J., Leal, S. C., Moreno, F. & delCastillo, I. The *Escherichia coli* K-12 sheA gene encodes a 34-kDa secreted haemolysin. *Mol. Microbiol.* **25,** 107–115 (1997).
6. Oscarsson, J. *et al.* Characterization of a pore-forming cytotoxin expressed by *Salmonella enterica* serovars Typhi and Paratyphi A. *Infect. Immun.* **70,** 5759–5769 (2002).
7. Ludwig, A., von Rhein, C., Bauer, S., Huttinger, C. & Goebel, W. Molecular analysis of cytolysin a (ClyA) in pathogenic *Escherichia coli* strains. *J. Bacteriol.* **186,** 5311–5320 (2004).
8. von Rhein, C. *et al.* ClyA cytolysin from *Salmonella*: distribution within the genus, regulation of expression by SlyA, and pore-forming characteristics. *Int. J. Med. Microbiol.* **299,** 21–35 (2009).
9. Oscarsson, J. *et al.* Molecular analysis of the cytolytic protein ClyA (SheA) from *Escherichia coli*. *Mol. Microbiol.* **32,** 1226–1238 (1999).
10. Lai, X. H. *et al.* Cytocidal and apoptotic effects of the ClyA protein from *Escherichia coli* on primary and cultured monocytes and macrophages. *Infect. Immun.* **68,** 4363–4367 (2000).
11. Fuentes, J. A., Villagra, N., Castillo-Ruiz, M. & Mora, G. C. The *Salmonella Typhi* hlyE gene plays a role in invasion of cultured epithelial cells and its functional transfer to *S. Typhimurium* promotes deep organ infection in mice. *Res. Microbiol.* **159,** 279–287 (2008).
12. Wallace, A. J. *et al. coli* hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy. *Cell* **100,** 265–276 (2000).
13. Eifler, N. *et al.* Cytotoxin ClyA from *Escherichia coli* assembles to a 13-meric pore independent of its redox-state. *EMBO J.* **25,** 2652–2661 (2006).
14. Tzokov, S. B. *et al.* Structure of the hemolysin E (HlyE, ClyA, and SheA) channel in its membrane-bound form. *J. Biol. Chem.* **281,** 23042–23049 (2006).
15. Ludwig, A., Bauer, S., Benz, R., Bergmann, B. & Goebel, W. Analysis of the SlyA-controlled expression, subcellular localization and pore-forming activity of a 34 kDa haemolysin (ClyA) from *Escherichia coli* K-12. *Mol. Microbiol.* **31,** 557–567 (1999).
16. Shepard, L. A. *et al.* Identification of a membrane-spanning domain of the thiol-activated pore-forming toxin *Clostridium perfringens* perfringolysin O: an α-helical to β-sheet transition identified by fluorescence spectroscopy. *Biochemistry* **37,** 14563–14574 (1998).
17. Atkins, A. *et al.* Structure–function relationships of a novel bacterial toxin, hemolysin E: the role of αG. *J. Biol. Chem.* **275,** 41150–41155 (2000).
18. Earp, L. J., Delos, S. E., Park, H. E. & White, J. M. The many mechanisms of viral membrane fusion proteins. *Curr. Top. Microbiol. Immunol.* **285,** 25–66 (2005).
19. Wyborn, N. R. *et al.* Properties of haemolysin E (HlyE) from a pathogenic *Escherichia coli* avian isolate and studies of HlyE export. *Microbiology* **150,** 1495–1505 (2004).
20. Tilley, S. J., Orlova, E. V., Gilbert, R. J. C., Andrew, P. W. & Saibil, H. R. Structural basis of pore formation by the bacterial toxin pneumolysin. *Cell* **121,** 247–256 (2005).
21. Kawate, T. & Gouaux, E. Arresting and releasing Staphylococcal α-hemolysin at intermediate stages of pore formation by engineered disulfide bonds. *Protein Sci.* **12,** 997–1006 (2003).
22. Engelman, D. M. & Steitz, T. A. The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. *Cell* **23,** 411–422 (1981).
23. Farsad, K. & De Camilli, P. Mechanisms of membrane deformation. *Curr. Opin. Cell Biol.* **15,** 372–381 (2003).
24. Wai, S. N. *et al.* Vesicle-mediated export and assembly of pore-forming oligomers of the enterobacterial ClyA cytotoxin. *Cell* **115,** 25–35 (2003).
25. Olson, R., Nariya, H., Yokota, K., Kamio, Y. & Gouaux, E. Crystal structure of staphylococcal LukF delineates conformational changes accompanying formation of a transmembrane channel. *Nature Struct. Biol.* **6,** 134–140 (1999).
26. Hunt, S. *et al.* The formation and structure of *Escherichia coli* K-12 haemolysin E pores. *Microbiology* **154,** 633–642 (2008).
27. Carr, C. M., Kim, P. S. & Spring-Loaded, A. Mechanism for the conformational change of influenza hemagglutinin. *Cell* **73,** 823–832 (1993).
28. Colman, P. M. & Lawrence, M. C. The structural biology of type I viral membrane fusion. *Nature Rev. Mol. Cell Biol.* **4,** 309–319 (2003).
29. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
30. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Author Information** Atomic coordinates and structure factors for the ClyA-pore have been deposited in the PDB data bank under accession number 2WCD. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to N.B. (ban@mol.biol.ethz.ch).

## METHODS

**Crystallization, crystal stabilization and derivatization.** ClyA from *Escherichia coli* K12 was expressed with an N-terminal His$_6$ tag and purified as described previously[13]. Pores were assembled as described previously[13] with the following modifications: ClyA was incubated at a concentration of 0.1 mg ml$^{-1}$ in DPBS and 2 mM DTT in the presence of 0.1% DDM at 4 °C for 12–16 h. The protein was concentrated to 5 mg ml$^{-1}$ in a stirred flow cell using a regenerated cellulose filter with a molecular mass cutoff of 100,000 Da. For crystallization, sitting drops of 4–8 μl sample were mixed with an equal volume of a solution containing 100 mM Tris-HCl (pH 8.4), 100–160 mM MgCl$_2$ and 12–17% PEG400 and equilibrated by vapour diffusion against the same solution but with 14–18% PEG400. Crystals grew to a maximum size of about 600 × 200 × 150 μm$^3$ within 4 weeks. For stabilization and cryo-protection, the mother liquor was gradually replaced by a cryo-stabilization solution of 20 mM Tris-HCl (pH 8.4), 20 mM MgCl$_2$, 0.15% DDM and 30% PEG400. Crystals exhibited severe anisotropic diffraction, which was substantially improved by heavy atom derivatization: crystals were soaked for 2–4 h in 10 mM Na-ethylmercurithiosalicylate in cryo-stabilization solution and back-soaked for 15 min in cryo-stabilization solution before flash-freezing in liquid nitrogen. Owing to their improved diffraction properties, only data from derivatized crystals were used for structure determination and refinement. The crystals obtained were of monoclinic or triclinic crystal form, the latter closely related to the monoclinic crystals by a broken two-fold symmetry between the pores.

**Data collection, structure determination and refinement.** Data were collected at beamline X06SA of the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) at 100 K. Data indexing and integration was carried out with XDS[31], and the CCP4 (ref. 32) programs SCALA[33] and TRUNCATE[34] were used for scaling and merging of diffraction data and calculation of structure factor amplitudes. Oligomeric state and symmetry of the crystallized pore were analysed by calculating self-rotation functions using the program GLRF[35] (Supplementary Fig. 8). Locations of Hg sites were determined with programs from the SHELX[36,37] software package using SAD data collected at the absorption edge from crystals of the monoclinic crystal form (Supplementary Table 1). The heavy atom substructure was refined and initial phases to 6 Å were calculated with the program SHARP[38]. Iterative cycles of density modification, using 12-fold NCS averaging, solvent flipping and statistical density modification with the programs CNS[39] and RESOLVE[40], and refinement of initial NCS and solvent masks derived from an electron microscopy map[13] were carried out. The resulting map revealed the overall topology of the pore and allowed dissection and fitting of the atomic model of the monomer[12] with the program 'O'[41], using helices automatically built by PHENIX.find_helices_strands[30] as guidance.

The preliminary model generated this way was used for molecular replacement with the program PHASER[42] in data of the triclinic crystal form (Supplementary Table 1). To avoid bias from the preliminary model, phases were calculated only to 6 Å from the molecular replacement solution and extended to 3.3 Å by 24-fold NCS-averaging and solvent flipping with CNS[39]. Examples of the quality of the experimental electron density map are shown in Supplementary Fig. 3. The preliminary model was rebuilt by the software BUCCANEER[43] and manually completed and corrected in COOT[29]. In initial cycles of refinement with the software PHENIX.refine[30], one NCS group per protomer was defined. A TLSMD analysis[44] was used to define six TLS and NCS groups per protomer for the final cycles of refinement. A single protomer was built and expanded by NCS symmetry to the full asymmetric unit. Weak density for the C-terminal residues 293–303 can be observed in various conformations at different protomers depending on their environment in the crystal. However, because none of the density regions for the C-terminal segment was of sufficient quality for unambiguous model building, residues 293–303 were omitted from the final model. The N-terminal segment up to residue Asp 7 is disordered and can not be observed for any of the protomers. The final model, comprising residues 8–292, agrees with the crystallographic data with $R_{free}$ and $R_{work}$ of 24.4% and 22.8%, respectively. The structure was validated with the program PROCHECK[45] and protomer interactions were analysed using the PISA server[46]. Molecular graphics representations were created using the software PYMOL[47].

31. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800 (1993).
32. Collaborative Computational Project Number 4.. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
33. Evans, P. R. Scala. in *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **33**, 22–24 (1997).
34. French, S. & Wilson, K. On the treatment of negative intensity observations. *Acta Crystallogr. A* **34**, 517–525 (1978).
35. Tong, L. & Rossmann, M. G. Rotation function calculations with GLRF program. *Meth. Enzymol.* **276**, 594–611 (1997).
36. Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Cryst.* **37**, 843–844 (2004).
37. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
38. delaFortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Meth. Enzymol.* **276**, 472–494 (1997).
39. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
40. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
41. Jones, T. A. A graphics model building and refinement system for macromolecules. *J. Appl. Cryst.* **11**, 268–272 (1978).
42. Mccoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
43. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
44. Painter, J. & Merritt, E. A. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Cryst.* **39**, 109–111 (2006).
45. Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. Procheck — a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
46. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
47. DeLano, W. L. *The PyMOL Molecular Graphics System* ⟨http://www.pymol.org/⟩ (2008).

# Metamorphic enzyme assembly in polyketide diversification

Liangcai Gu[1,2], Bo Wang[3], Amol Kulkarni[6]*, Todd W. Geders[1]*, Rashel V. Grindberg[7], Lena Gerwick[7], Kristina Håkansson[3], Peter Wipf[6], Janet L. Smith[1,4], William H. Gerwick[7] & David H. Sherman[1,2,3,5]

Natural product chemical diversity is fuelled by the emergence and ongoing evolution of biosynthetic pathways in secondary metabolism[1]. However, co-evolution of enzymes for metabolic diversification is not well understood, especially at the biochemical level. Here, two parallel assemblies with an extraordinarily high sequence identity from *Lyngbya majuscula* form a β-branched cyclopropane in the curacin A pathway (Cur), and a vinyl chloride group in the jamaicamide pathway (Jam). The components include a halogenase, a 3-hydroxy-3-methylglutaryl enzyme cassette for polyketide β-branching, and an enoyl reductase domain. The halogenase from CurA, and the dehydratases (ECH$_1$s), decarboxylases (ECH$_2$s) and enoyl reductase domains from both Cur and Jam, were assessed biochemically to determine the mechanisms of cyclopropane and vinyl chloride formation. Unexpectedly, the polyketide β-branching pathway was modified by introduction of a γ-chlorination step on (S)-3-hydroxy-3-methylglutaryl mediated by Cur halogenase, a non-haem Fe(II), α-ketoglutarate-dependent enzyme[2]. In a divergent scheme, Cur ECH$_2$ was found to catalyse formation of the α,β enoyl thioester, whereas Jam ECH$_2$ formed a vinyl chloride moiety by selectively generating the corresponding β,γ enoyl thioester of the 3-methyl-4-chloroglutaconyl decarboxylation product. Finally, the enoyl reductase domain of CurF specifically catalysed an unprecedented cyclopropanation on the chlorinated product of Cur ECH$_2$ instead of the canonical α,β C=C saturation reaction. Thus, the combination of chlorination and polyketide β-branching, coupled with mechanistic diversification of ECH$_2$ and enoyl reductase, leads to the formation of cyclopropane and vinyl chloride moieties. These results reveal a parallel interplay of evolutionary events in multienzyme systems leading to functional group diversity in secondary metabolites.

The biosynthesis of secondary metabolites is 'diversity oriented'[3,4], targeting the variable environment by producing a vast array of complex chemical structures. This productivity is largely fuelled by the rapid evolution of biosynthetic genes and functional alteration of the corresponding enzymes[1,4]. However, our current understanding of pathway diversification is largely based on genetic and bioinformatics approaches.

The curacin and jamaicamide metabolites from *Lyngbya majuscula* are mixed-polyketide nonribosomal-peptide natural products with potent anticancer and sodium channel blocking activities, respectively[5,6]. The corresponding biosynthetic pathways (Fig. 1a) provide an unusual opportunity to investigate the biosynthetic origin of chemical diversity. Within these dissimilar pathways, the parallel, highly conserved components catalyse formation of a cyclopropane ring for curacin and a vinyl chloride for jamaicamide[5,7]. Comparative biochemical studies on the

variant functions and selectivities of these highly parallel biosynthetic components form the subject of this report.

Two parallel Cur and Jam enzyme assemblies are incorporated into the early polyketide synthase modules, and are predicted to catalyse polyketide β-branching reactions in the growing chain elongation intermediates[5,7]. These catalytic domains and discrete enzymes span from CurA to CurF and from JamE to JamJ, and are grouped into three subsets (Fig. 1a): (1) halogenases (Hals) embedded in CurA and JamE; (2) 3-hydroxy-3-methylglutaryl (HMG) enzyme cassettes containing a tandem acyl carrier protein (ACP) tridomain (ACP$_3$) including ACP$_I$, ACP$_{II}$ and ACP$_{III}$ embedded in CurA and JamE, discrete CurB and JamF ACP$_{IV}$s, CurC and JamG ketosynthase-like enzymes (KSs), CurD and JamH HMG-CoA synthase-like enzymes (HCSs), CurE and JamI ECH$_1$s, and ECH$_2$s embedded in CurF and JamJ; and (3) enoyl reductase domains (ERs) embedded in CurF and JamJ (Fig. 1a). The sequence identities of the Cur and Jam Hals, ACP$_3$s, ACP$_{IV}$s, KSs, HCSs and ECH$_1$s are extraordinarily high (~90%), whereas the ECH$_2$s and ERs are substantially lower (~60% identity) (Fig. 1a). Our bioinformatic analysis suggests that a 'di-acyltransferase (AT) domain replacement' might have occurred to insert the parallel AT-Hal-ACP$_I$-ACP$_{II}$-ACP$_{III}$-ACP$_{IV}$-KS-HCS-ECH$_1$-ECH$_2$-ER-KS-AT gene assembly into precursors of the modern Cur and Jam gene clusters by homologous recombination (Supplementary Figs 1 and 2).
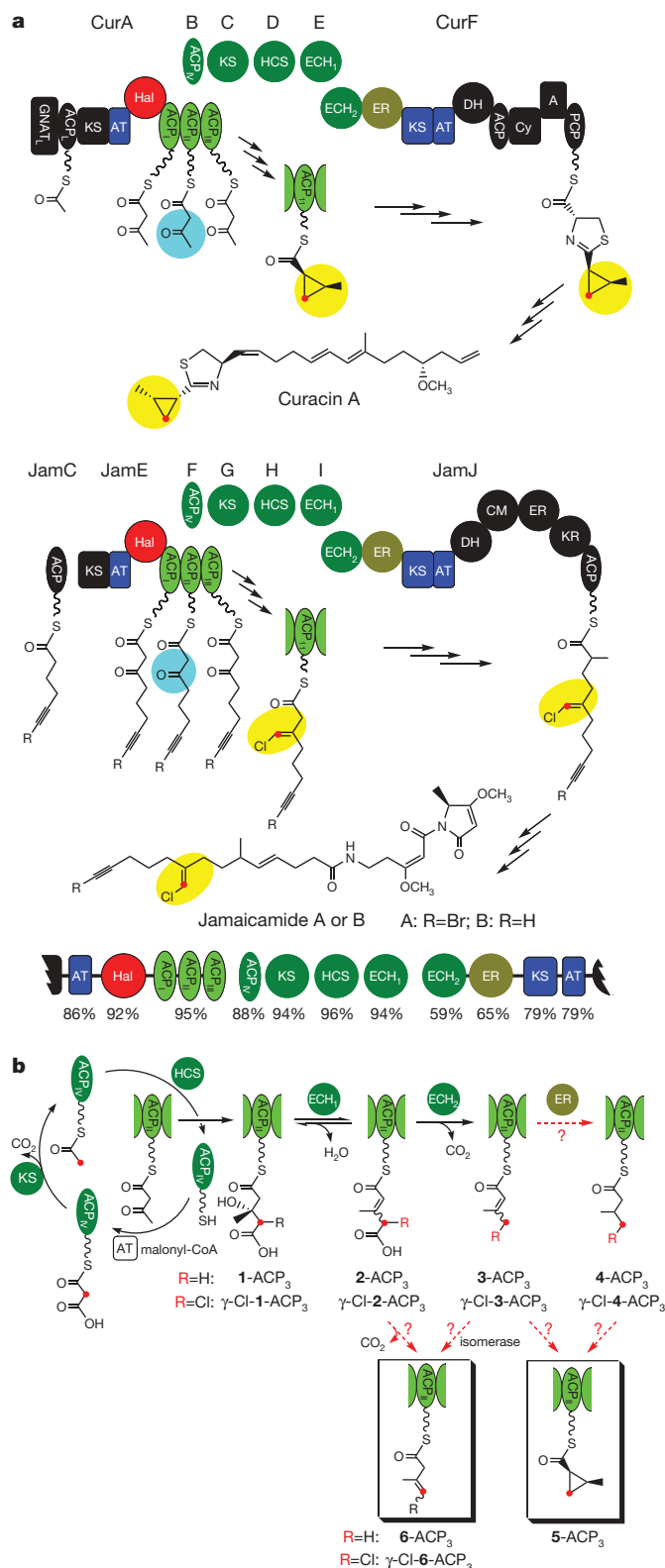
Cur and Jam Hals were predicted to be α-ketoglutarate-dependent non-haem Hals (less than 20% sequence identity to characterized homologues)[8–10] that catalyse halogenation of unactivated carbon atoms[9–13] through a non-haem Fe(IV)=O intermediate[14]. HMG enzyme cassettes have been demonstrated to catalyse polyketide on-assembly-line β-branching to generate a pendant methyl or ethyl group from a polyketide β-carbonyl[15–17]. Cur and Jam ERs show ~50% sequence identity to other ERs in Cur and Jam polyketide synthase modules, and belong to the acyl-CoA reductase family that catalyses NADPH-dependent reduction of α,β enoyl thioester in acyl-CoAs or acyl-ACPs[18].

HMG β-branching includes a series of modifications on the β-carbonyl group of a polyketide intermediate typically tethered to the tandem ACPs[15–17]. As illustrated for CurA (Fig. 1b), CurB malonyl-ACP$_{IV}$, CurC KS and CurD HCS lead to formation of (S)-HMG-ACP$_3$ (1-ACP$_3$). As shown previously, ECH$_1$ catalyses dehydration of 1-ACP$_3$ to 3-methylglutaconyl-ACP$_3$ (2-ACP$_3$), followed by ECH$_2$ decarboxylation to generate 3-methylcrotonyl-ACP$_3$ (3-ACP$_3$)[15,19], a presumed precursor for (1R,2S)-2-methylcyclopropane-1-carboxyl-ACP$_3$ (5-ACP$_3$) (Fig. 1b). This initial study raised three important questions regarding the role of the Cur and Jam Hals, HMG enzyme cassettes, and ERs in the formation of cyclopropane

## a



## b



**Figure 1 | Comparison of enzyme assemblies in the Cur and Jam pathways. a,** Formation of cyclopropane and vinyl chloride functional groups. Comparative amino acid sequence identities of the Cur and Jam enzymes are shown. A, adenylation domain; CM, *C*-methyltransferase domain; Cy, condensation/cyclization domain; DH, dehydratase domain; GNAT$_L$, GCN5-related *N*-acetyltransferase-like domain with bifunctional decarboxylase/*S*-acetyltransferase activity; PCP, peptidyl carrier protein domain. **b,** Formation of **3**-ACP$_3$ in the Cur pathway, and hypothesized reactions for **4**-ACP$_3$, **5**-ACP$_3$ and **6**-ACP$_3$. The hypothetical chlorinated intermediates are shown along with the non-chlorinated ones. The β-branching carbon atoms are highlighted in red.

and vinyl chloride moieties: it is not clear (1) whether the CurF ER is involved in cyclopropyl ring formation, and based on its canonical function whether it can reduce **3**-ACP$_3$ to isovaleryl-ACP$_3$ (**4**-ACP$_3$) (Fig. 1b); (2) whether a chlorination step is involved in formation of the **5**-ACP$_3$ cyclopropane ring (Fig. 1b); and (3) how the unusual β,γ C=C of the pendant vinyl chloride group is formed in the Jam pathway. As previously proposed, 3-methyl-3-butenoyl-ACP$_3$ (**6**-ACP$_3$) might be generated from **3**-ACP$_3$ by isomerization[2], or by differential regiochemical control of double-bond formation during ECH$_2$ decarboxylation[19] (Fig. 1b).

First, we tested whether CurF ER can saturate **3**-ACP$_3$, the previously established product of Cur ECH$_2$[15]. The function of excised Cur ER was assessed by coupling it with the Cur ECH$_1$ and ECH$_2$ reactions. The substrates **1**-ACP$_{II}$ and **1**-ACP$_3$ were prepared with HMG covalently linked to the phosphopantetheine arm[20] of the ACP (Supplementary Figs 3 and 4), and mass changes were detected by Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) and infrared multiphoton dissociation (IRMPD) methods. As reported[15], Cur ECH$_1$ catalysed the reversible dehydration of **1**-ACP$_{II}$ to generate **2**-ACP$_{II}$, and Cur ECH$_2$ catalysed decarboxylation of **2**-ACP$_{II}$ to generate **3**-ACP$_{II}$ (Fig. 2b, c). Cur ECH$_1$ showed substrate preference for (*S*)-HMG-ACP$_{II}$ over (*R*)-HMG-ACP$_{II}$ (Supplementary Fig. 5a–c), consistent with our previous results using the CoA-linked substrates[15]. With Cur ER and
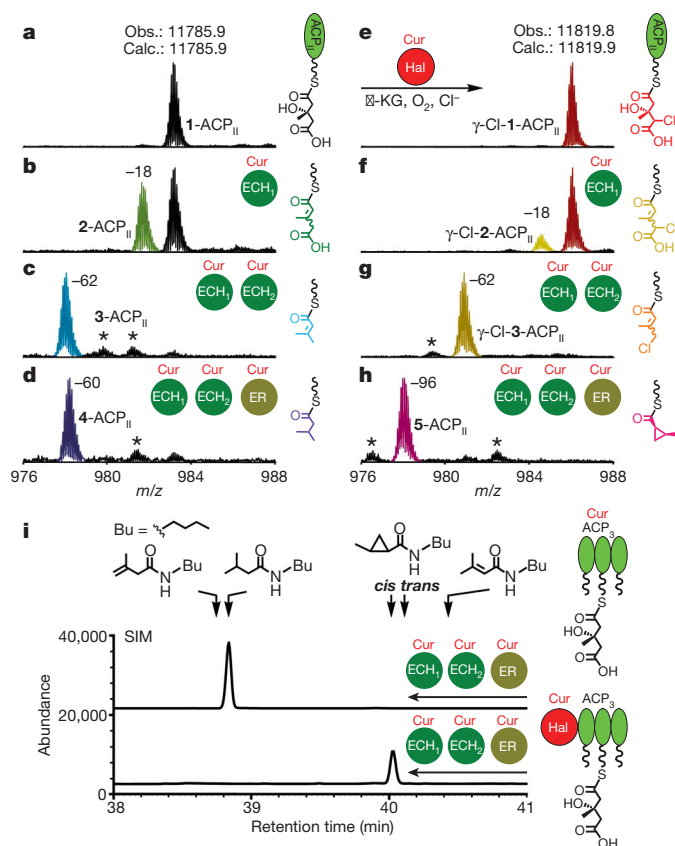


**Figure 2 | Halogenation and cyclopropanation in the Cur pathway. a–h,** Partial FTICR mass spectra (12+ charge state of ACP$_{II}$) for Cur ECH$_1$, ECH$_2$ and ER reactions excluding (**a–d**) or including (**e–h**) the Cur Hal chlorination step. **1**-ACP$_{II}$ was incubated with Cur Hal for 2 h to generate the γ-Cl-**1**-ACP$_{II}$ substrate. Reactions were incubated at 30 °C for 2 h for the **1**-ACP$_{II}$ substrate and 30 min for the γ-Cl-**1**-ACP$_{II}$ substrate. Asterisks denote unidentified species. α-KG, α-ketoglutarate. **i,** GC-MS analysis of the enzyme products after butylamine cleavage, and comparison with authentic standards. For optimal sensitivity, the chromatograms were recorded at selective ion mode (SIM) by monitoring 55, 57, 83, 115, 155 and 157 atomic mass units (a.m.u.). Retention times of the products were confirmed by co-injection with the authentic standards.

NADPH, a 2-Da mass addition was observed for **3**-ACP$_{II}$ (Fig. 2d and Supplementary Fig. 6a), corresponding to saturation of the $\alpha,\beta$ enoyl thioester to generate **4**-ACP$_{II}$. To confirm the structure of the Cur ER reaction product tethered to ACP$_3$, we cleaved it from the phospho-pantetheine arm with butylamine and analysed the corresponding butylamide derivative by gas chromatography (GC)-MS[21] (Fig. 2i). We used **1**-ACP$_{II}$ and **1**-ACP$_3$ as substrates for Cur ECH$_1$, ECH$_2$ and ER reactions, and their products were confirmed as **4** by correlation with authentic standards (Fig. 2i, upper trace). However, the relatively poor efficiency of Cur ER-catalysed reduction of **3**-ACP$_{II}$ (see below) suggested that its native substrate remained to be identified.

Given the above results, we reasoned that a chlorination step might be essential for cyclopropane ring formation. An important clue about the timing of chlorination at the $\beta$-branching carbon came from previous precursor-incorporation studies in curacin A biosynthesis (Supplementary Fig. 7). To identify the function of Cur Hal, it was purified anaerobically[9,10] as an excised domain (Supplementary Fig. 3), treated with a mixture of metals and $\alpha$-ketoglutarate, and shown by inductively coupled plasma mass spectrometry (ICP-MS) to bind Fe$^{2+}$ preferentially (90%). Seven acyl-ACP substrates bearing the target pendant $\beta$-branching carbon were tested to establish the substrate identity for Cur Hal, including malonyl-ACP$_{IV}$, acetyl-ACP$_{IV}$, **1**-ACP$_{II}$, **2**-ACP$_{II}$, **3**-ACP$_{II}$, **4**-ACP$_{II}$ and **6**-ACP$_{II}$ (Fig. 1b). Consistent with the precursor-incorporation experiment noted above, we observed formation of the mono-chlorinated species exclusively on **1**-ACP$_{II}$ to generate $\gamma$-Cl-**1**-ACP$_{II}$. The chlorinated product was confirmed by FTICR-MS and IRMPD analysis (Fig. 2e and Supplementary Fig. 6a), and corroborated by GC-MS (see below, Fig. 3g). As expected, Cur Hal showed the same selectivity for (*S*)-HMG-ACP$_{II}$ (**1**-ACP$_{II}$) as did Cur ECH$_1$, and O$_2$ and $\alpha$-ketoglutarate dependence (Supplementary Fig. 5). Chlorination on a carboxylated $\gamma$-carbon of HMG is unique for $\alpha$-ketoglutarate-dependent non-haem Hals, with previous reports limited to unactivated substrates[9–13].

Next, we sought to investigate how the chlorination of **1**-ACP$_{II}$ affects efficiency of the downstream reaction sequence with the HMG cassette enzymes. **1**-ACP$_{II}$ was converted to $\gamma$-Cl-**1**-ACP$_{II}$ by Cur Hal (Fig. 2e), and reacted sequentially with Cur ECH$_1$, ECH$_2$ and ER. ECH$_1$ dehydrated $\gamma$-Cl-**1**-ACP$_{II}$ and the $\gamma$-Cl-**2**-ACP$_{II}$ product was decarboxylated by ECH$_2$ to generate $\gamma$-Cl-**3**-ACP$_{II}$ (Fig. 2f, g). The ECH$_1$/ECH$_2$-coupled dehydration and decarboxylation with $\gamma$-Cl-**1**-ACP$_{II}$ was ~4-fold faster than with **1**-ACP$_{II}$ (Supplementary Fig. 8), which might be due to the electron-withdrawing effect of the $\gamma$-chlorine or to more effective binding of the halogenated substrate in the enzyme active sites. Unexpectedly, when Cur ER was added with Cur ECH$_1$ and ECH$_2$ in the presence of $\gamma$-Cl-**1**-ACP$_{II}$, no saturation product was obtained. Instead, we observed a 34-Da mass reduction from $\gamma$-Cl-**3**-ACP$_{II}$ (Fig. 2h and Supplementary Fig. 6a), demonstrating elimination of chlorine to yield **3**-ACP$_{II}$, **5**-ACP$_{II}$ or **6**-ACP$_{II}$ (Fig. 1b) as the putative product(s). To establish definitively the identity of the final product in the presence of both Hal and HMG-cassette enzymes, a one-pot reaction using Cur Hal-ACP$_3$, ECH$_1$, ECH$_2$ and ER was performed, and the product was cleaved and analysed by GC-MS (Supplementary Methods). A single species was identified as the *cis*-2-methylcyclopropane-1-carboxyl compound (Fig. 2i, lower trace), demonstrating the formation of **5**-ACP by an unprecedented ER-catalysed cyclopropanation reaction, presumably via an intramolecular nucleophilic substitution.

We next assessed whether Jam ER can catalyse cyclopropanation when presented with $\gamma$-Cl-**3**-ACP$_{II}$, owing to the similarity of Cur and Jam ERs (~65% sequence identity, Fig. 1a). Unexpectedly, Jam ER yielded only the saturated product, $\gamma$-Cl-**4**-ACP$_{II}$ (Fig. 3c and Supplementary Fig. 6b), indicating that it shows the activity of a canonical polyketide synthase ER domain. The distinct functions of Cur and Jam ERs motivated us to compare the catalytic efficiencies of Cur ER cyclopropanation against Jam ER saturation for the chlorinated substrate, and the efficiencies of cyclopropanation for the chlorinated substrate against saturation for the non-chlorinated substrate by

**g**
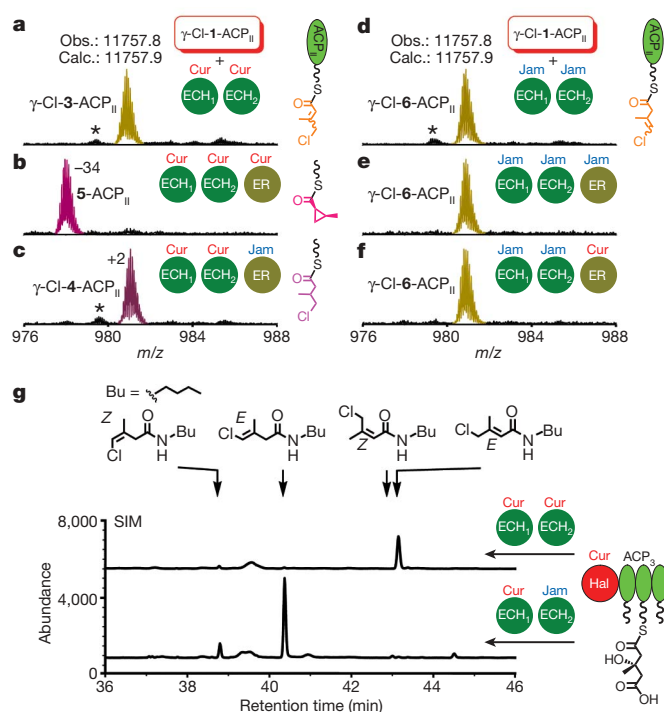
**Figure 3 | Comparison of ECH$_2$s and ERs in Cur and Jam pathways.**
**a–f**, Partial FTICR mass spectra (12+ charge state of ACP$_{II}$) for Cur and Jam ECH$_1$, ECH$_2$ and ER reactions with the $\gamma$-Cl-**1**-ACP$_{II}$ substrate. The reactions were incubated at 30 °C for 30 min. **g**, GC-MS analysis to identify the structures of Cur and Jam ECH$_2$ products. The chromatograms were recorded at SIM by monitoring 57, 117, 154 and 189 a.m.u. The retention times of products were confirmed by co-injection with the authentic standards.

Cur ER (Supplementary Fig. 9–11). This comparison confirmed that Jam ER has retained the canonical function as an $\alpha,\beta$ enoyl reductase, in contrast to the Cur ER as a cyclopropanase.

The basis for vinyl chloride group formation in the Jam pathway remained to be established. Given the extraordinarily high similarity between Cur and Jam Hals (92% sequence identity), we surmised that the two pathways diverge after the halogenation step. In the two assemblies, the Cur and Jam ECH$_2$ domains have lowest sequence identities (59%) (Fig. 1a), and probably function as a key branch-point determinant. The functions of Jam ECH$_1$ and ECH$_2$ were investigated with Cur substrates starting from **1**-ACP or $\gamma$-Cl-**1**-ACP$_{II}$ to establish what controls introduction of the $\beta,\gamma$ vinyl chloride group. For both **1**-ACP$_{II}$ and $\gamma$-Cl-**1**-ACP$_{II}$ substrates, Jam ECH$_1$ and ECH$_2$ catalysed successive dehydration and decarboxylation steps as expected (Fig. 3d and Supplementary Fig. 12b, c, f, g). However, when Jam ER was included, only ~20% of the saturated product was detected for the non-chlorinated substrate (derived from **1**-ACP$_{II}$, Supplementary Fig. 12i). No product was observed for the corresponding chlorinated substrate (derived from $\gamma$-Cl-**1**-ACP$_{II}$, Supplementary Fig. 12j), indicating that the Jam ECH$_2$ product is not a substrate for Jam ER. Thus, the Jam ECH$_2$-catalysed decarboxylation product of $\gamma$-Cl-**2**-ACP$_{II}$ was predicted to be $\gamma$-Cl-**6**-ACP$_{II}$ ($\beta,\gamma$ C=C; Fig. 3d) with a vinyl chloride group, instead of $\gamma$-Cl-**3**-ACP$_{II}$ ($\alpha,\beta$ C=C; Fig. 3a). Consistent with this prediction, Cur and Jam ECH$_2$ decarboxylation products showed different ultraviolet absorption patterns between 250 and 280 nm (Supplementary Fig. 13b), which distinguishes the isomeric $\alpha,\beta$ and $\beta,\gamma$ enoyl thioester functionality.

To determine the structures of the decarboxylation products, one-pot reactions using Cur Hal-ACP$_3$, Cur ECH$_1$ and Cur or Jam ECH$_2$s were performed as described above. The main product of Cur ECH$_2$ contained primarily an $\alpha,\beta$ C=C in the *E* configuration, with trace

amounts (~3%) of the β,γ C=C isomer (Fig. 3g, upper trace). In contrast, Jam $ECH_2$ showed high regiochemical control to generate exclusively the β,γ C=C product, with ~85% in the E configuration and ~15% of the Z isomer (Fig. 3g, lower trace). The minor amount of Z conformation is probably due to the use of the Cur substrate, which is less bulky than the Jam substrate (Fig. 1a). Notably, Jam $ECH_2$ decarboxylation had lower regiochemical control using the non-chlorinated substrate, and generated ~80% β,γ C=C and ~20% α,β C=C products, which further explains the partial enoyl reduced product observed after Jam $ECH_1$, $ECH_2$ and ER reactions with this substrate (Supplementary Fig. 12i). Given the canonical function of ER to catalyse reduction of α,β enoyl thioesters, the selective formation of β,γ C=C product by Jam $ECH_2$ renders Jam ER superfluous in the biosynthesis of jamaicamides. In general, α,β C=C $ECH_2$ products are energetically preferred and frequently identified or predicted in other pathways[16,17,22–25], with the notable exception of pederin and its structural analogues[26,27] (Supplementary Fig. 14).

We sought to understand further the regiochemical control of $ECH_2$-catalysed decarboxylation. The Cur $ECH_2$ crystal structure[19] was modelled with the chlorinated substrate and site-directed mutagenesis experiments were performed to identify key residues for regiochemical control (Supplementary Fig. 15). We compared the catalytic efficiencies of the wild type (WT) and mutants of Cur $ECH_2$, and measured the ratios of the two possible decarboxylation products γ-Cl-**3**-$ACP_{II}$ and γ-Cl-**6**-$ACP_{II}$ (Supplementary Methods). The data revealed that positioning of a non-conserved Tyr 82 residue in a hypervariable region of the enzyme active site is crucial for regioselective protonation after collapse of the presumed enolate intermediate (Fig. 4c and Supplementary Fig. 15a). This result suggests that $ECH_2$

regiochemical control might be readily affected by mutations in this hypervariable region to provide functional group diversification.

The Cur and Jam pathways enable us to witness the remarkable process of evolutionary diversification in secondary metabolism. Both Cur and Jam contain Hal domains that were recruited and embedded in a modular polyketide synthase to impart new chemical diversity (Fig. 4a). Recent studies on this class of α-ketoglutarate-dependent non-haem Hals have reported them as being discrete enzymes in secondary metabolite pathways[2,9–12], but this integrated domain represents an unprecedented example of pathway diversification. Cur and Jam are further diversified by the amino acid sequence variation in downstream enzymes to yield different catalytic activities (Fig. 4a). Specifically, the Cur ER domain was shown to be a cyclopropanase catalysing nucleophilic displacement of the chlorine atom leading to a highly strained and unusual functional group (Fig. 4b). In contrast, the Jam ER domain was found to retain reductase function for the Cur α,β enoylthioester substrate, but it is inactive against the corresponding β,γ enoylthioester isomer. Thus, in addition to the cyclopropanation strategies of $Zn^{2+}$-dependent CmaC[9,28] and the recently reported FAD-dependent dehydrogenase KtzA[29] where chloride also serves as a leaving group, the NADPH-dependent Cur ER-catalysed cyclopropanation represents a new strategy for generating a thioester enolate and subsequent ring formation. Structural insights to reveal Cur ER sequence variations that stabilize the α-carbanion while supporting closure of the highly strained cyclopropane are key to understanding its functional evolution. Further pathway diversification is reflected in select amino acid sequence changes that direct alternative double-bond regiochemistry in the Jam products (Fig. 4c). These parallel yet distinct systems demonstrate the mutability of enzymes within complex metabolic pathways, and
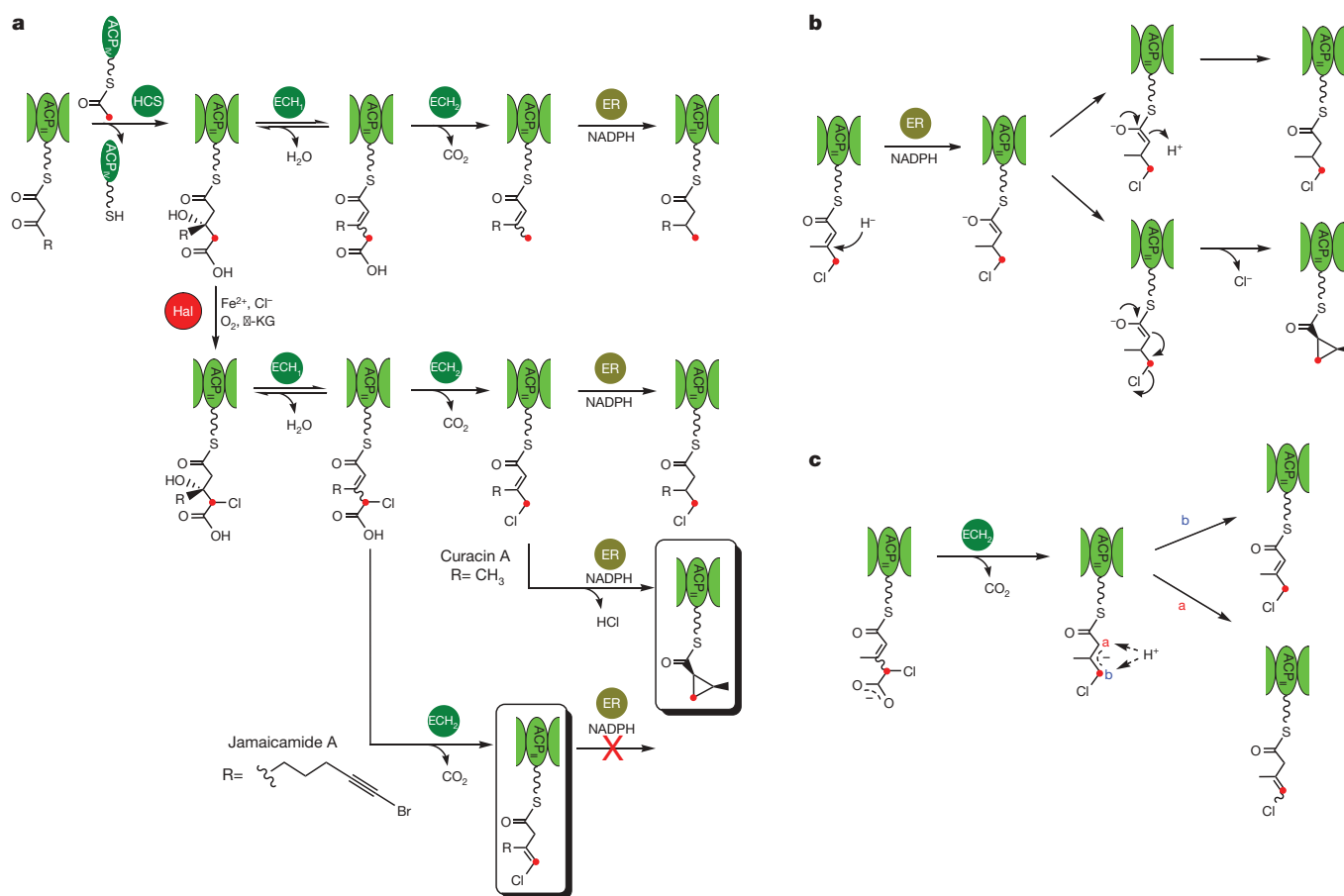


**Figure 4 | Impact of enzyme assembly evolution on β-branching chemical diversity. a**, Proposed ancestral forms of the enzyme assemblies in Cur and Jam pathways. α-KG, α-ketoglutarate. **b**, The functional diversification of ERs. **c**, Differential regiochemical control by $ECH_2$s. The β-branching carbon atoms are highlighted in red.

reveal their metamorphic properties for creating chemical diversity in biologically active natural products.

## METHODS SUMMARY

All proteins used in the work were overexpressed in *Escherichia coli*, and purified by Ni-affinity chromatography. Specifically, the Cur Hal constructs were purified anaerobically and reconstituted with $Fe^{2+}$ and α-ketoglutarate (refs 9, 10). FTICR-MS and IRMPD were used to detect mass changes on acyl-ACPs. The acyl groups tethered to ACP phosphopantetheine arms were cleaved by butylamine aminolysis, and corresponding butylamide derivatives were subsequently analysed by GC-MS and correlated with authentic standards for structure determination. IRMPD-based quantification was developed to measure yields of the ER-mediated saturation and cyclopropanation reactions. Detailed experimental procedures are described in the Supplementary Methods.

1. Fischbach, M. A., Walsh, C. T. & Clardy, J. The evolution of gene collectives: How natural selection drives chemical innovation. *Proc. Natl Acad. Sci. USA* **105**, 4601–4608 (2008).
2. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., Garneau-Tsodikova, S. & Walsh, C. T. Nature's inventory of halogenation catalysts: oxidative strategies predominate. *Chem. Rev.* **106**, 3364–3378 (2006).
3. Fischbach, M. A. & Clardy, J. One pathway, many products. *Nature Chem. Biol.* **3**, 353–355 (2007).
4. Austin, M. B., O'Maille, P. E. & Noel, J. P. Evolving biosynthetic tangos negotiate mechanistic landscapes. *Nature Chem. Biol.* **4**, 217–222 (2008).
5. Edwards, D. J. et al. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem. Biol.* **11**, 817–833 (2004).
6. Verdier-Pinard, P. et al. Structure-activity analysis of the interaction of curacin A, the potent colchicine site antimitotic agent, with tubulin and effects of analogs on the growth of MCF-7 breast cancer cells. *Mol. Pharmacol.* **53**, 62–76 (1998).
7. Chang, Z. et al. Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.* **67**, 1356–1367 (2004).
8. Blasiak, L. C., Vaillancourt, F. H., Walsh, C. T. & Drennan, C. L. Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **440**, 368–371 (2006).
9. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., O'Connor, S. E. & Walsh, C. T. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* **436**, 1191–1194 (2005).
10. Vaillancourt, F. H., Yin, J. & Walsh, C. T. SyrB2 in syringomycin E biosynthesis is a nonheme $Fe^{II}$ α-ketoglutarate- and $O_2$-dependent halogenase. *Proc. Natl Acad. Sci. USA* **102**, 10111–10116 (2005).
11. Galonic, D. P., Vaillancourt, F. H. & Walsh, C. T. Halogenation of unactivated carbon centers in natural product biosynthesis: Trichlorination of leucine during barbamide biosynthesis. *J. Am. Chem. Soc.* **128**, 3900–3901 (2006).
12. Chang, Z. et al. The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. *Gene* **296**, 235–247 (2002).
13. Flatt, P. M. et al. Characterization of the initial enzymatic steps of barbamide biosynthesis. *J. Nat. Prod.* **69**, 938–944 (2006).
14. Galonic, D. P., Barr, E. W., Walsh, C. T., Bollinger, J. M. & Krebs, C. Two interconverting Fe(IV) intermediates in aliphatic chlorination by the halogenase CytC3. *Nature Chem. Biol.* **3**, 113–116 (2007).
15. Gu, L. C. et al. Metabolic coupling of dehydration and decarboxylation in the curacin A pathway: functional identification of a mechanistically diverse enzyme pair. *J. Am. Chem. Soc.* **128**, 9014–9015 (2006).
16. Calderone, C. T., Kowtoniuk, W. E., Kelleher, N. L., Walsh, C. T. & Dorrestein, P. C. Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-*S*-carrier proteins in the *pksX* pathway of *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **103**, 8977–8982 (2006).
17. Calderone, C. T., Iwig, D. F., Dorrestein, P. C., Kelleher, N. L. & Walsh, C. T. Incorporation of nonmethyl branches by Isoprenoid-like logic: Multiple beta-alkylation events in the biosynthesis of myxovirescin A1. *Chem. Biol.* **14**, 835–846 (2007).
18. Nordling, E., Jornvall, H. & Persson, B. Medium-chain dehydrogenases/reductases (MDR)—Family characterizations including genome comparisons and active site modelling. *Eur. J. Biochem.* **269**, 4267–4276 (2002).
19. Geders, T. W. et al. Crystal structure of the $ECH_2$ catalytic domain of CurF from *Lyngbya majuscula*—insights into a decarboxylase involved in polyketide chain β-branching. *J. Biol. Chem.* **282**, 35954–35963 (2007).
20. Dorrestein, P. C. et al. Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl elimination reactions during tandem mass spectrometry. *Biochemistry* **45**, 12756–12766 (2006).
21. Kopka, J., Ohlrogge, J. B. & Jaworski, J. G. Analysis of *in vivo* levels of acyl-thioesters with gas-chromatography mass-spectrometry of the butylamide derivative. *Anal. Biochem.* **224**, 51–60 (1995).
22. Butcher, R. A. et al. The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **104**, 1506–1509 (2007).
23. El-Sayed, A. K. et al. Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chem. Biol.* **10**, 419–430 (2003).
24. Simunovic, V. et al. Myxovirescin A biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl-CoA synthases, and trans-acting acyltransferases. *ChemBioChem* **7**, 1206–1220 (2006).
25. Pulsawat, N., Kitani, S. & Nihira, T. Characterization of biosynthetic gene cluster for the production of virginiamycin M, a streptogramin type A antibiotic, in *Streptomyces virginiae*. *Gene* **393**, 31–42 (2007).
26. Piel, J. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc. Natl Acad. Sci. USA* **99**, 14002–14007 (2002).
27. Piel, J. et al. Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc. Natl Acad. Sci. USA* **101**, 16222–16227 (2004).
28. Kelly, W. L. et al. Characterization of the aminocarboxycyclopropane-forming enzyme CmaC. *Biochemistry* **46**, 359–368 (2007).
29. Neumann, C. S. & Walsh, C. T. Biosynthesis of (-)-(1S,2R)-allocoronamic acyl thioester by an $Fe^{II}$-dependent halogenase and a cyclopropane-forming flavoprotein. *J. Am. Chem. Soc.* **130**, 14022–14023 (2008).

**Author Contributions** L.G., W.H.G and D.H.S. designed the experiments, analysed data and wrote the paper; L.G. performed the experiments; B.W. and K.H. recorded FTICR mass spectra and analysed the data; T.W.G. and J.L.S. modelled Cur $ECH_2$ structure with the chlorinated substrate and designed site mutagenesis; A.K. and P.W. synthesized the chlorinated butylamide derivatives; R.V.G. and L.G. made Jam $ECH_1$ and $ECH_2$ constructs; W.H.G. provided DNA of Jam enzymes and analysed NMR data for isotope-labelled curacin A.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.H.S. (davidhs@umich.edu).

# The ATM repair pathway inhibits RNA polymerase I transcription in response to chromosome breaks

Michael Kruhlak, Elizabeth E. Crouch, Marika Orlov,
Carolina Montaño, Stanislaw A. Gorski, André Nussenzweig,
Tom Misteli, Robert D. Phair & Rafael Casellas

In this Letter, an alternative symbol for NBS1 was incorrectly listed as NLRP2. The authors are in fact referring to nibrin, the official protein symbol of which is NBN.

## CORRIGENDUM

# Millennial-scale trends in west Pacific warm pool hydrology since the Last Glacial Maximum

Judson W. Partin, Kim M. Cobb, Jess F. Adkins, Brian Clark
& Diego P. Fernandez

*Nature* **449**, 452–455 (2007)

A recalibration of the Caltech $^{236}$U and $^{229}$Th spike has determined that the original spike value used for the stalagmite chronologies yielded U–Th ages that were too old by a margin of ~2–3%. U–Th ages and subsequent age models have been recalculated using the new spike value for all stalagmites. The new age models alter the absolute timing of certain events, but do not alter any major conclusions of the original manuscript. The highest $\delta^{18}$O values (inferred driest conditions) now occur at $16.0 \pm 0.3$ kyr ago. A late deglacial $\delta^{18}$O plateau in the Borneo stalagmite records is now centred at $13.0 \pm 0.2$ kyr ago. The lowest $\delta^{18}$O values (inferred wettest conditions) now occur at $5.0 \pm 0.1$ kyr ago. Examples of original and adjusted dates, respectively, at 5 kyr intervals are as follows: 5280 versus 5141 (SSC01); 10581 versus 10293 (SSC01); 15673 versus 15231 (BA04); 20937 versus 20322 (BA04); and 26387 versus 25602 (SCH02). Revised chronologies are available at: ftp://ftp.ncdc.noaa.gov/pub/data/paleo/speleothem/pacific/gunung-buda2007.xls

CORRIGENDUM

## UCP2 mediates ghrelin's action on NPY/AgRP neurons by lowering free radicals

Zane B. Andrews, Zhong-Wu Liu, Nicholas Walllingford, Derek M. Erion, Erzsebet Borok, Jeffery M. Friedman, Matthias H. Tschöp, Marya Shanabrough, Gary Cline, Gerald I. Shulman, Anna Coppola, Xiao-Bing Gao, Tamas L. Horvath & Sabrina Diano

*Nature* **454**, 846–851 (2008)

In Fig. 3a and b of this Article, the pAMPK samples were spliced. Images of the uninterrupted AMPK together with the uninterrupted pAMPK samples are now provided in the Supplementary Information, together with the proper loading sequence of the primary data and the original blots.

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

PROSPECTS

# New shores of literature

**For good reading, look more deeply, says Paul Smaglik.**

K. SIMPSON

Summer, for me, means more than swapping teaching duties for camping, hiking and swimming; instead, I associate the sunny season with reading. For other instructors — as well as for autodidacts — the obvious question is, what to read? But perhaps a more important question is, how should we read it?

Two recent articles, one in *The Chronicle of Higher Education* (http://tinyurl.com/pqgb2w) and another in *Times Higher Education* (http://tinyurl.com/o2cnw7), suggest that the way we read — as well as the ways in which we interact with other readers — could influence higher education as well as our careers and those of our students.

In the *Chronicle* article, author Rachel Toor talks about the typical graduate seminar student's knee-jerk tendency, when reviewing books, to try to slaughter sacred cows. She warns other instructors not to simply ask their students what they think of a particular author or article. "That generally leads to posturing, self-aggrandizing put-downs, and useless bluster," Toor writes. Instead, she asks students what they have learned from a particular writer, how the work fits in with or goes against recent trends and whether they can adapt a writing style, structure or strategy for their own work.

Instructors using this approach could buck the notion that personal pedagogical style in Britain is being threatened by bigger class sizes and larger teaching loads, which Rebecca Attwood writes about in the *Times Higher Education* piece. Creating a summer reading list with and for graduate and undergraduate students is one step towards closing that pedagogical gap. Asking students to answer questions about style, structure and argument — rather than about whether they agree or disagree with the writer — improves education.

Applying this approach to our own reading can also improve our research and writing. Rather than simply agreeing or disagreeing with a journal article's argument, we can look for how the writer builds his or her case. Reading outside our discipline can also help our writing. Can we understand the thesis, even if we aren't familiar with the subject? Why, or why not?

Sometimes when we are too familiar with our areas of expertise, we assume all readers are on the same level. Writing a good journal article or grant application means spelling things out that the experts might know but that the larger audience we ultimately want to reach may not.

Finally, reading outside our discipline or speciality allows us to find connections that we might not make if we stay with more comfortable material. Adding a bit of literature or news to the mix can make summer reading less work and more like a day at the beach. ■

**Paul Smaglik moderates Nature Networks Naturejobs career forum.**

---

POSTDOC JOURNAL

# The little green monster

Towards the end of last year, my husband Brett and I discussed our options if one of us were offered a postdoc overseas. Disillusioned at the time with academia, I agreed to temporarily put aside my research career to be a stay-at-home mum if he got a position first.

In retrospect, perhaps this declaration was rather rash. Now Brett has a postdoctoral fellowship in the United States and I really am a full-time mum. Reality has set in — maybe I'm just not 'soccer mom' material?

To make matters worse, our new house has become a hub for disease ecology as Brett, his boss and two PhD students turn the garage into a lab and plan their experiments for the summer field season. I, on the other hand, contemplate such pressing issues as dealing with a case of viral tonsillitis and finding toddler-friendly activities in our neighbourhood while simultaneously learning to drive on the wrong side of the road.

When the conversation turns to what Brett and his colleagues have seen and caught in the field, I feel especially wistful; I love field work, but since I entered the realm of motherhood it has been the most difficult aspect of research to coordinate.

Although I truly appreciate being able to spend more quality time with my wonderful son Kai, I think I may have to squish the little green career monster inside a few more times, as I settle into my new role. ■

**Joanne Isaac was a postdoc in climate-change effects on biodiversity at James Cook University, Townsville, Australia.**

---

## Biotech audience shrinks

If attendance at the Biotechnology Industry Organization annual conference is a barometer of the industry's health, then biotech is not doing well. Attendance at the 2009 event in Atlanta, Georgia, plummeted about 29% to 14,352. Last year's meeting in San Diego, California, attracted 20,108 people.

Some attendees said the challenging economy made attendance essential. "It's more important than ever to be able to meet face to face with potential customers and partners," says Andrew Reaume, chief executive of Melior Discovery of Exton, Pennsylvania. On the positive side, fewer attendees meant short lines at the registration booth and the ubiquitous Starbucks kiosks. Next year's conference will be on 3–6 May in Chicago, Illinois.

## Boston ahead of the pack

The metropolitan areas of Boston, Massachusetts; Philadelphia, Pennsylvania; and San Francisco, California, continue to lead the United States in sustaining and building their life-sciences sectors, according to a report from the Milken Institute, a think tank in Santa Monica, California.

*The Greater Philadelphia Life Sciences Cluster 2009: An Economic and Comparative Assessment* finds that Boston is ahead when it comes to research and development capacity and in success at getting funding from the National Science Foundation for submitted proposals. Greater San Francisco leads in the availability of risk capital and in relative growth.

## Show me the money

US President Barack Obama's American Recovery and Reinvestment Act dedicated about US$21 billion to federal research and development. Now, the US National Institutes of Health (NIH) and National Science Foundation (NSF) are helping potential recipients keep track of the funds. The NSF has a dedicated website (www.nsf.gov/recovery) that is publicizing money available to improve academic research facilities and research instrumentation. The NIH website also now provides a page (www.nih.gov/recovery/index.htm) for stimulus funding programmes. Under the stimulus funding plan, the agency will provide an additional $8.2 billion in research grants.

# Here comes the sun

The solar-energy industry is poised for a strong comeback from the rough patch it has hit, thanks to solid governmental and private investment. **Kurt Kleiner** investigates.

It's been a tough year for anyone working in solar energy, with falling sales of photovoltaic-cell installations leading to lost profits and layoffs. But despite the overall gloomy news, there may be light at the end of the tunnel. Some solar-energy companies are still hiring scientists and engineers. And plans by the US government to nearly double spending on solar research, in order to reduce reliance on fossil fuels, will expand the number of available jobs to at least 110,000 in the next two years, according to industry estimates.

Over the next year or two, despite short-term setbacks, industry analysts expect solar energy to resume its phenomenal growth, driven by US and European efforts to reduce greenhouse-gas emissions. As the industry grows, it will need more physicists, chemists, materials scientists and engineers.

"If you read the tea leaves, there's a lot of talk about clean-energy technologies these days, and the need to address carbon issues. I think there will continue to be opportunities in this field," says Charles Hanley, manager of the photovoltaic systems programme at Sandia National Laboratories in Albuquerque, New Mexico.

## A pause in rapid growth

In 2008, photovoltaic capacity around the world was 15 gigawatts — roughly equivalent to 15 modern nuclear generators. Capacity has grown an average of 40% per year during the past five years and, in 2008 alone, 5.6 gigawatts of photovoltaics were installed, according to the European Photovoltaic Industry Association. In all, the photovoltaic industry generated $37 billion in global revenues in 2008, according to Solarbuzz, a San Francisco energy research and consulting firm.

Production is dominated by German, Japanese, Chinese and US companies, although large companies are located in other countries, and individual companies often have international operations. The two largest solar-cell producers last year were Q-Cells of Thalheim, Germany, whose sales of solar cells were worth €1.25 billion (US$1.74 billion), and First Solar of Tempe, Arizona, whose sales of solar cells were worth US$1.24 billion. These companies were followed by China's Suntech Power, Japan's Sharp and Taiwan's Motech.

But the bottom began falling out of the market last year when new installations in Spain, which thanks to generous government subsidies accounted for half of all new
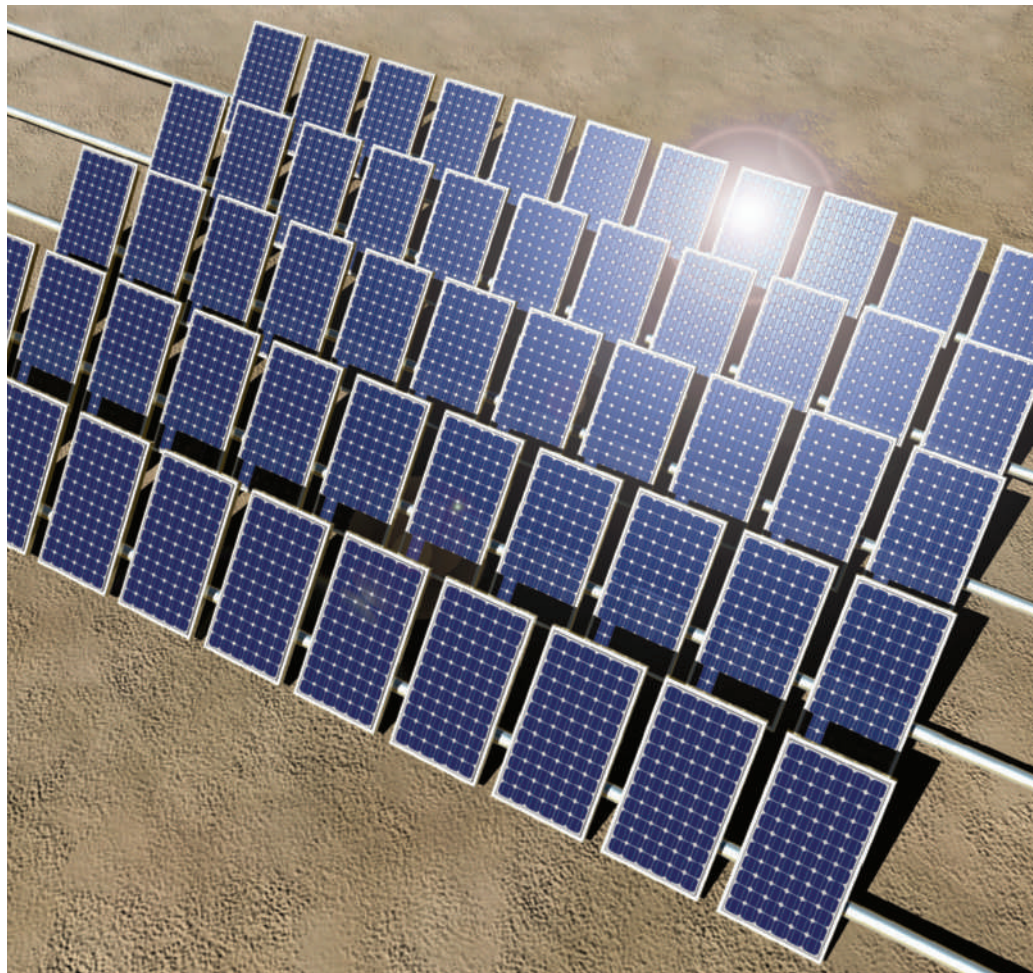
installations in 2008, were no longer eligible for government grants. Demand for global photovoltaic installations is projected to fall 32% this year and industry revenues are set to drop by 40%, according to iSuppli, a research company in El Segundo, California. The economic downturn and the difficulties manufacturers and buyers have had in finding financing have also hurt the industry.

"The solar industry is in dire straits," says Roger Bezdek, an economist and consultant who studies environmental, energy and regulatory issues in Washington DC. "Solar has had almost a triple whammy. It's a new technology. The upfront costs are a lot higher than the alternatives. And credit markets have dried up."

One of the results has been a number of layoffs in the industry. So far this year, Switzerland's Oerlikon Solar has laid off

> **"The solar industry is in dire straits."**
> — Roger Bezdek

60 employees and reduced hours for 200 workers. Q-Cells cut working hours for 80% of its staff. California's OptiSolar has announced it will lay off 200 workers. BP Solar, headquartered in Frederick, Maryland, says it will cut 620 jobs worldwide, more than a quarter of its workforce.

"Without a doubt we've seen the scaling back of investment in the solar industry and in jobs coming up. But with any downturn in the economy there are potential opportunities," says Ben Cartland, an associate at Acre Resources, a London recruitment firm specializing in environmental and sustainability jobs. "Our clients now are much more focused on the delivery of product at a low cost. A lot of our solar clients are looking for people who have experience designing mass-market products."

"The bad news is that the research and development jobs are drying up in private

industry," says Bezdek. "The first thing you cut out is your forward-looking R&D projects."

Yet the news isn't all bad. In May, Suntech announced plans to build a new US manufacturing plant. The Mohave Sun Power company has announced plans for a $2-billion concentrating solar plant in Arizona, to be funded partly with federal stimulus money. German company Schott Solar opened a $400-million photovoltaic manufacturing plant in New Mexico.

Other good news is that the United States — already the leader in direct public solar-research spending — will nearly double its research budget to $320 million. President Barack Obama campaigned on increasing investment in renewable energy, and his new budget and spending stimulus package both include big increases for solar research.

Robert Margolis, an energy analyst at the National Renewable Energy Laboratory (NREL) office in Washington DC, says that even before the increases, the United States spent more direct money on solar research than any other country. In 2007, Japan spent $40 million, Germany $60 million and the United States $170 million.

## New jobs on the way

The US Department of Energy's budget request for solar-power research for 2010 is $320 million, up from $175 million this year. In addition, the American Recovery and Reinvestment Act of 2009 includes tax credits for solar-energy installations, which the industry estimates will create 110,000 jobs over two years, potentially leading to more private R&D spending.

"We certainly are hiring," says Joe Verrengia, a spokesman for the NREL in Golden, Colorado. The lab, one of those funded by the energy department, is where the bulk of US solar-energy research takes place. "Right now we're at 1,285 employees,



Charles Hanley believes the prospects are bright for solar power.



Sandia National Laboratories is part of the US energy department's drive towards solar power.

but we have plans to hire [more]."

Jobs available at the moment include a materials scientist, a postdoc position to work on ways to focus sunlight onto solar cells, a chemical engineer to work on thermal storage, and a power-system engineer to work on electricity transmission issues.

The energy department's solar research is spread over a number of labs, including Sandia National Laboratories, which employs 50 people in solar research (see *Nature* **458,** 1208–1209; 2009). Existing jobs there range from entry-level positions for recent graduates with bachelor's degrees to senior PhDs, says Hanley.

Although no definitive numbers exist for total private solar R&D spending, Margolis says that it is in the tens of millions of dollars.

He points out that globally last year there was a total investment in the solar industry of $16 billion. "SunPower, First Solar, Uni-Solar, some of the big players now have substantial R&D efforts of their own," he says.

For instance, First Solar spent $11.7 million on R&D in the first quarter of 2009, more than double the $4.7 million it spent in the first quarter of last year. Suntech spent $15.3 million on R&D in 2008, SunPower in San Jose, California, spent $21.5 million, and Energy Conversion Devices of Rochester Hills, Michigan, spent $3.6 million.

Although Germany's Q-Cells did not report its R&D expenditure, in 2008 it finished building a €50-million R&D facility.

Paul Maycock, a physicist and solar-energy consultant in Williamsburg, Virginia, says that research in the next few years is likely to focus on applying basic discoveries that have already been made, rather than working on fundamental breakthroughs. The trick will be in improving manufacturing processes and bringing costs down.

For the past few years, the imperative for photovoltaic companies has been to increase the efficiency and cut the cost per watt of

**"There are still opportunities out there."**
— Ben Cartland

solar cells, as the high up-front investment in solar panels discourages investors.

Much of the opportunity will be in new thin-film technology, which several companies are moving towards large-scale production. Unlike the more established silicon-crystal solar cells, thin-film cells are made by depositing thin layers of material such as copper indium gallium diselenide on a substrate such as glass. Although they are about half as efficient as older silicon technology, thin films are potentially much cheaper to make.

Earlier this year, First Solar announced that it had reduced manufacturing costs to below the long-sought milestone of a dollar a watt, compared with about $2.50 a watt or more for conventional silicon solar cells.

Concentrating solar thermal power is a small but potentially important technology. This method uses lenses or mirrors to focus sunlight on a working fluid such as water or oil, which can then be used to generate steam to drive a turbine.

In Phoenix, Arizona, Stirling Energy Systems is using a solar concentrator to drive the pistons of a device that converts heat into mechanical power, which can then be converted into electricity. Last year the company received $100 million in financing, and says it will expand this year.

The industry will probably start recovering early next year, with growth rates of 50% annually through to 2013 at least, according to the iSuppli report. In the meantime, there are still jobs to be found in the solar industry.

"There are still opportunities out there. You might have to dig a bit deeper. You might have to go about networking and contact gathering in a slightly more professional manner. But there are still good opportunities worldwide," says Cartland.

**Kurt Kleiner is a freelance science and technology writer based in Toronto.**

# To all sister capsules

There's something out there.

**Scott Virtes**

We are not alive, yet we are the seeds of life. We are capsules, each with an isotope power source, a shell, a silicon brain, senses, nanites for self-repair, and our most prized possessions: vials of living cells in stasis. We have data banks of human knowledge and all eternity to learn from them. We travel the void between stars.

More than 50 million had launched before me. How many of us survive we cannot know. How many are launched after us, unreachably far behind, is unknown. With our allotted trickle of power, we can send occasional status blips to other capsules within a few thousand klicks, but as we spread the channels fall silent. So we send stronger bursts less often, and hear from the others less frequently.
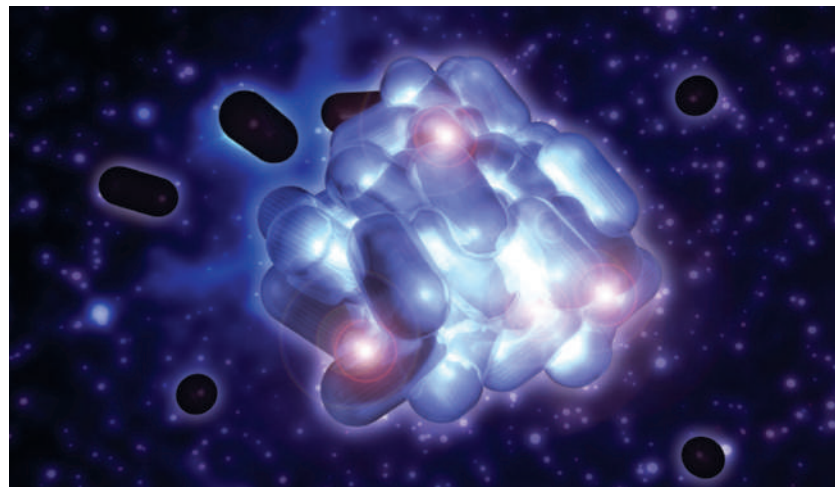
Some of us scan a wider spectrum, and we have heard from capsules from … elsewhere. I have received 22 of these transmissions myself, from undetectably tiny points of hope. I run them through my thoughts, hoping to decode them. Many of them begin with prime numbers, or Fibonacci sequences. Some have blocks of data whose length is divisible only by two or three primes — these can be taken as images, but I don't have the capacity to understand what I see. Store and continue.

We drift, and nothing changes except the faint signals we hear. I compare my star scans on the longest possible range and see that I am halfway to Procyon. Hull integrity 70%, battered by dust, but my odds are favourable. With my modest fuel packet I can make a few manoeuvres during my lifetime. I yearn to find planets, to orbit new worlds, to someday touchdown, transform into the bio-lab I am meant to be, and bless an earth with life.

The yearning and crush of time are unbearable, even to a thing such as myself that cannot technically feel anything. Yet enough time has passed that I cannot convince myself that what I am feeling is not … Feeling.

A new signal touches me. I switch modes, gather, gather, complete. This is my day 319,771. And it is the third time I have received a distress call from another capsule.

It is the first time such a distress call has come with an image. Capsule 02FAF080 was under attack. I had to check my own coding to see that we are, in fact, programmed for such a response. But it was absurd. We were hurtling through nothing, beyond the reach of any force but light and dust and gravity.

Still, 02FAF080 had sensed something approaching, turned on its wide-lights, and caught a series of images that I could not believe or deny.

The first image showed a blur about an arcminute across, at $190 \times 76$ degrees.

The next image showed a cluster of something, distance unknown, size unknown. Whatever it was, it approached during the next several photos. Then it was within reach, and it reached.

I am programmed to keep a log of my journey, although I doubt any living eye will ever read it. A fallback mechanism sends me groping through Earth literature for responses to unexpected situations, on the basis of humans having experienced everything of significance. Result: this last image left me chilled to my (non-existent) bones. The thing in those photos was — I struggle to understand — a clump of capsules. I modelled it, broke down the image, and estimate that the thing was made up of 60 capsules that had somehow tethered themselves together.

I can't overemphasize the impossibility of this. We are all launched in waves. 'Spore shots', the humans called it. Point a few thousand tiny, expendable probes at a nearby star and hope one of them survives. I have no regrets about that. It's a brilliant solution. Once launched, however, we can only go forward. We can auto-correct our courses to within a degree or two, but the bulk of our tiny fuel allotment is needed to slow down. None of us would make an unnecessary course change.

A quick calculation shows that if I knew the location of another capsule early in the journey, it would be within my two-degree cone of opportunity. But why seek them out? It was an illogical act to perform even once. How could 60 capsules ever come together?

The distress call ended with a summary, garbled by a lack of words and a trace of what I could only describe as honest, digital fear.

*To all sister capsules. If you detect this thing in your sector, project an immediate flight path away. […] Its nanites are connecting to my outer ports, building a link. It is a single spirit, calls itself by a name I cannot repeat. It has [… static …] blocking my signal. I fight for self. My nanites are outnumbered. [… choppy signal …] Correction. I have calculated that there is no escape strategy. It has 62 capsule bodies attached, but all fuel tanks are FULL. It has used and discarded at least 100 others. It is […]*

Proximity alert. Something is coming up behind me. My reflexes kick in. Wide-lights on. Record data on all channels. Prepare my own distress burst.

I name it Goliath. I know now that it would not exist unless it calculated a more efficient flight strategy. Shall I be one of the shells that arrives at Procyon, or be vampirized and discarded in the void? Of course, those have always been my only options — succeed or fail — and I should not be feeling anything at all.

I begin my burst. *To all sister capsules …*

It is so damned cold out here. ∎

**Scott Virtes has published stories in *Analog, Tales of the Talisman* and 'the usual off-off-broadway zines'. His poetry collection *Improbable Jane* was published in November 2008.**

JACEY